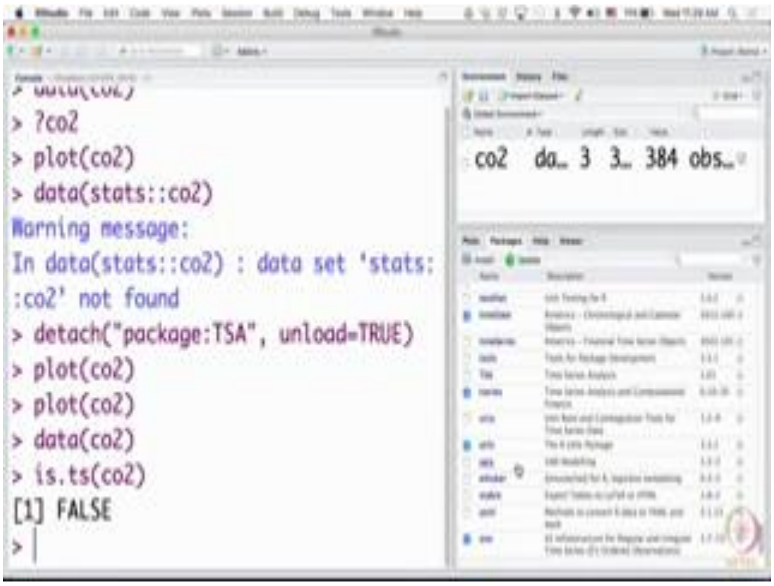
Applied Time-Series Analysis
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture – 54
Lecture 23B - Models for Linear Non-stationary Processes 2

What I will do is, of course there are other methods we will come back to that, I just want to show you how you would fit trends in r using LM or the another package. So, let me go back now to r.

(Refer Slide Time: 00:33)



Increase the font size to Rayleigh font, now what we are going to do is let me actually here throw away all the variables here they are not gone thank you. So, let us start a fresh session here, one of the data sets that exhibits strengths is this very nice data set which is the data set is nice, but the phenomenon is not so great this is a carbon dioxide emission.

So, what we can do is let us actually load this data set, it comes pre installed with your base packages it is the co2; it may be good for example, to read up something on the data set of course, is asking which co2 you are loading, we can actually load this in the I think it is loading from the TSA. So, it is a level of carbon dioxide that alerts Canada, I think it is loading that so that is what it has done for me. If you if you are not sure is there are other there is another co2 here, but I think this is the z this is from a different geographical region.

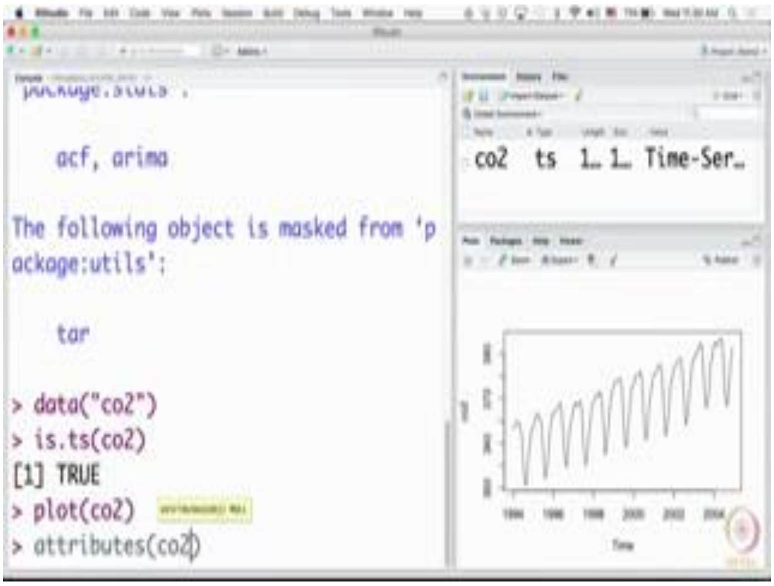
So, let us plot and see whether it has the features that we want; this is not the one that we want it is it has a trend and some oscillations no worries. So, let us work with this data and see the other one that is that the co2 that is present in the a (Refer Time: 02:24) this well and do it oops let us see if it works (Refer Time: 02:28) oops not found or maybe I can just unload the TSA just to show you that it is a different data set, but that also as a trend this is one of the confusions in our. So, I am going to actually unload is the same thing or maybe because co2 is already be loaded.

So, this is a different co2 this also as the features that we want it has this linear trend I do not know how are you going to see you are able to see let me zoom out. So, this is the trend you can see that it as this trend and then there are the oscillatory features and then maybe it has some randomness in it, we are not able to see that at least visually if; in the classical approach you think of this series as being made up of trend plus some periodic component plus a random component.

Now, as I said there is a non parametric approach that one can adopt and a parametric approach first of all this is a time series object, when we load co2, I think you should be able to see that let us see better close, this is a data frame you can see it is not a time series object it is a data frame.

On the other hand if you have to go back and load the reload the TSA package.

(Refer Slide Time: 04:18)

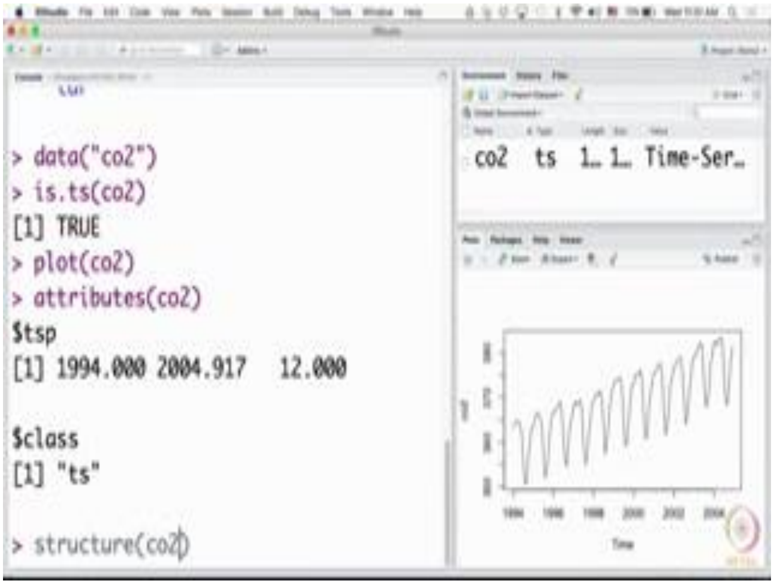


And then get the co2 from the TSA package, you get a nice time series object of course you can always convert a data frame object to a TSA objects it is not difficult. So, let us ask now I am going back to this co2 in the TSA package and it clearly shows it is a time series object, if you are not able to see that there you go it is says true and it has also the similar trend.

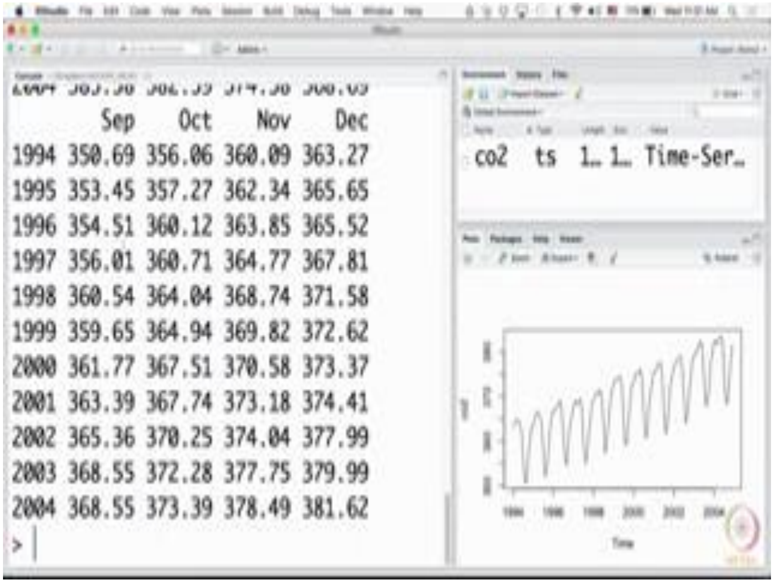
But now this carbon dioxide emission has been recorded in a different job difficult region over a different period of time. So, let us say we work with these; this also has a trend and a periodic component and so on. Now let us say I want to fit a linear trend to this, the classic approach is to use the in r is to use the l m routine, where you fit a model of this form alpha series as alpha naught and alpha 1 k.

Now, if you have to do that you have to first extract or construct the times series vector because the regressor; so the time stamp factor sorry the regressor here is a time instant itself. Now that can take some time first of all it may be good to know when this was recorded. So, it was recorded in this period as you can see and it is a monthly co2 data how do I know that well I can ask for the structure.

(Refer Slide Time: 05:55)



(Refer Slide Time: 06:00)



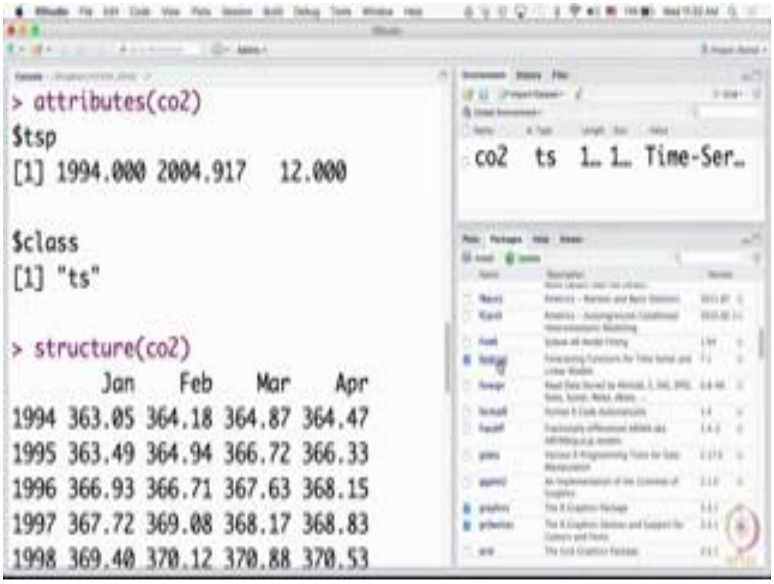
As you can see it has been recorded on a monthly basis from Jan to December all months and over these years from 1994 sorry to 2004.

We say that this is actually recorded on a yearly basis, but with a frequency of 12. So, the unit of observation is a year you can think of it that way, but you are within a unit of observation you are observing 12 times. So, that is why when you ask for the frequency which is shown here when you look at the attributes of the time series object, the third one the first is the start point, second is the stop point and the third is the frequency with which it has been measured and it has been measured 12 times in a year that is what it is telling you.

Now, as I said if you want to fit a trend of this form we have to construct a time vector and that can be a bit painful it is not difficult, but it can be a bit cumbersome you might want a routine which looks into the time series object straightaway takes the data and the time stamps and fix up a trend for you using the least squares method, and that is what that is what the tslm from the forecast package, you should install a forecast package it is a fantastic package it has a lot of nice routines in it.

One of the routines that you will really love is a routine that automatically estimates the order of the ARIMA model for you, but do not get pampered by that routine be careful it does help lot of people say that they did finite useful, but they were also misled right; any automated method for order determination should also be used with caution right.

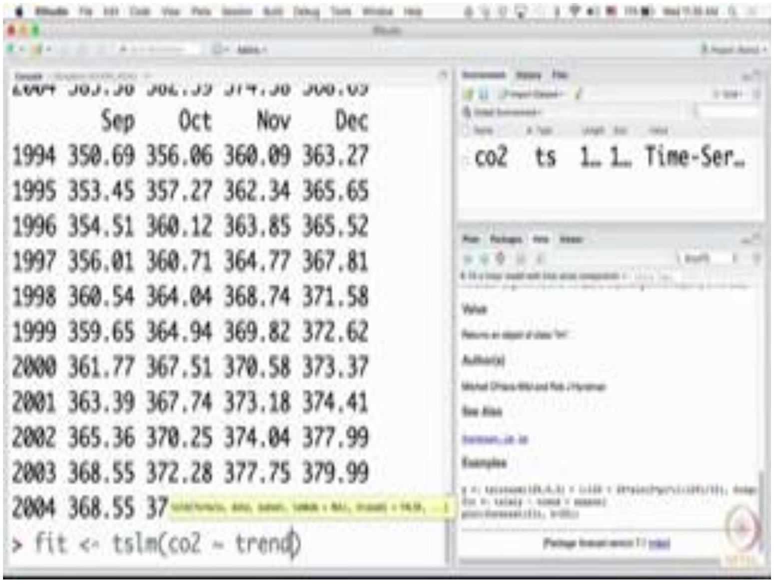
(Refer Slide Time: 08:00)



So, I have load a installed a forecast package and it is also loaded to my r works with a environment. So, there you go this is the forecast; the forecast package is a number of useful routines which we will use later on as well. So, let us actually use the tslm In fact, if you go to the routines available in the forecast package and pull up the help on t s l m towards the end of this help file there is a simple example which generates a synthetic time series it is has a trend plus sinusoid and a periodic signal sorry and your random component. So, you can see I do not know how well you can see at the bottom here.

But there is a first component is a regular one and then now you have the sinusoid component and then you have a trend as well here. So, this is the first one is your random component here r naught and then you have the linear trend followed by a sinusoids. So, that is put together you can work through that example as well, but just let us look at the co2 data here.

(Refer Slide Time: 09:14)

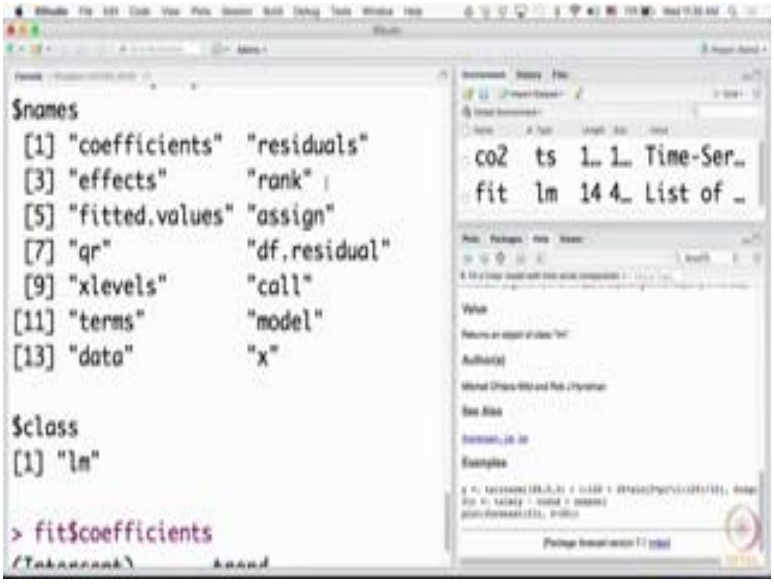


Now, what we are going to do is, we are going to fit at linear trend and obtain the results in a variable called fit. So, t s l m and the nice thing is you can say here co2, the formula here if you recall I have shown you the syntax of the formal l m, you use the tilde operated to express the relationship; the co2 is a series for us and we want to fit a trend and here it is assumed that the trend when you say trend it is linear trend whether it is going to fit.

And whereas, if you were to use the l m you have to explicitly specify the regressor, where the regressor here is k of course, you also have this constant that is automatically fit by l m for you. So, regressor here is k with l m you have to construct the time stamp vector and pass it on as a regressor, here you do not have to do that just ask it to fit a trend and it will do it for you.

So it has done it and it is important to know what attributes this object has.

(Refer Slide Time: 10:30)

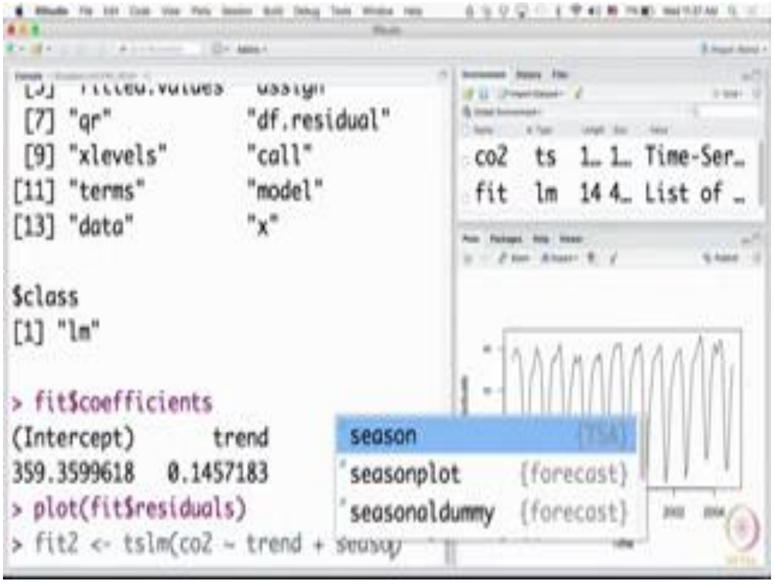


So, it says that these are all the things that are stored in the fit the coefficients alpha naught and alpha 1 for example, we can ask for that. So, there you go the intercept term and the trend right alpha naught is the first one and alpha 1 is the second they are estimates of that and what we are interested in is of course, the residuals right we want to see if the linear trend has sufficed for this series.

So, let us plot the residuals, there are other attributes as well, but we will not worry about that now what do you think, have you managed to get rid of the trend why, but then what else do you see now as expected you see a periodic component? Now, I am going to talk about a generic approach to this in general there are other routines.

For example there is something called STL a routine called STL, which comes with the stats package, it stand for seasonal trend decomposition and it uses a method called loess, but it asks the user for example, to specify the seasonality, for that you should have some information a priori even in this I could have asked for example, to fit both the seasonal component and the trend component in other words the periodic and the trend.

(Refer Slide Time: 12:09)



So, what do you expect to see if it has done a good job of fitting the seasonal component? You expect to see some random component like this right there is no particular trend or seasonality and so on.

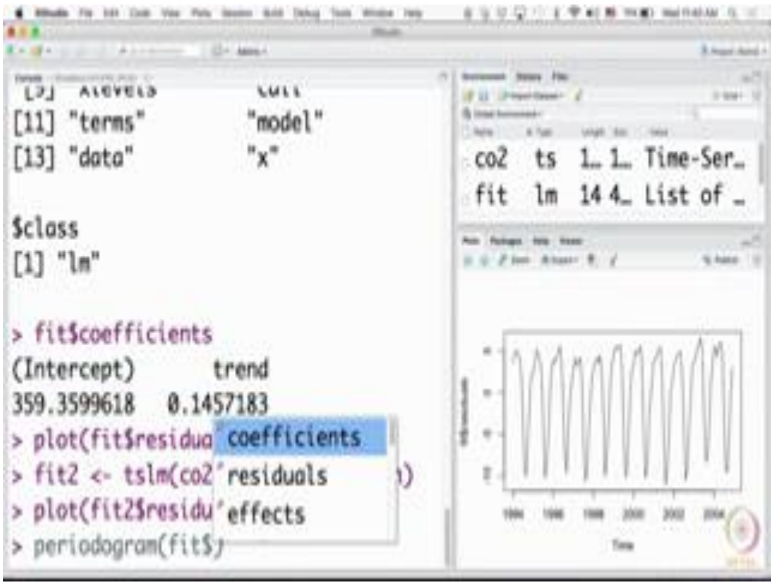
It has automatically done that for you, but when you have set season what it has done is it has assumed that the underlying component as a periodicity of 12; that is you remember we said we are measuring we have 12 observations in a year so it assumes that the periodicity, I have strictly speaking the observation interval is one month. So, it by saying that the frequency is 12 it reads that frequency and it assumes that the periodicity is 12 of course, you can specify you can overwrite that and say no do not assume 12 there is some other periodicity as well.

But how do you know what the periodicity is? that is the big question that generally goes that is unknown a priori in certain cases you may know it, but in a lot of other cases you do not know the periodicity, one has to discover it through a data analysis and that is where your spectral analysis comes handy, which we will talk introduce very soon this week.

So, the spectral analysis allows you to detect periodicities and one of the ways to detect periodicity is of course, to look at the power spectrum of the series and figure out what is where the peaks are just to give you a preview we will go over the theory, but a lot of you must be familiar with power spectrum; a power spectrum plot is a plot of power versus frequency and we search for peaks in this spectral plot, at whatever frequency you notice a peak you say a well that is a component is present predominantly.

Now, we will learn later a measure or a concept known as periodogram, which we will do this for you a periodogram is nothing, but a power spectral density plot empirical power spectral density plot, which plot power versus frequency and there is a routine called spectrum in the stats package, but I am not so fond of that because it plots not the power, but decibels rather 20 log 10 of power I am not so comfortable with decibels and so on. So, I do not like the plot that it produces and then you have to go through additional steps to make sure it plots only the absolute one rather than that.

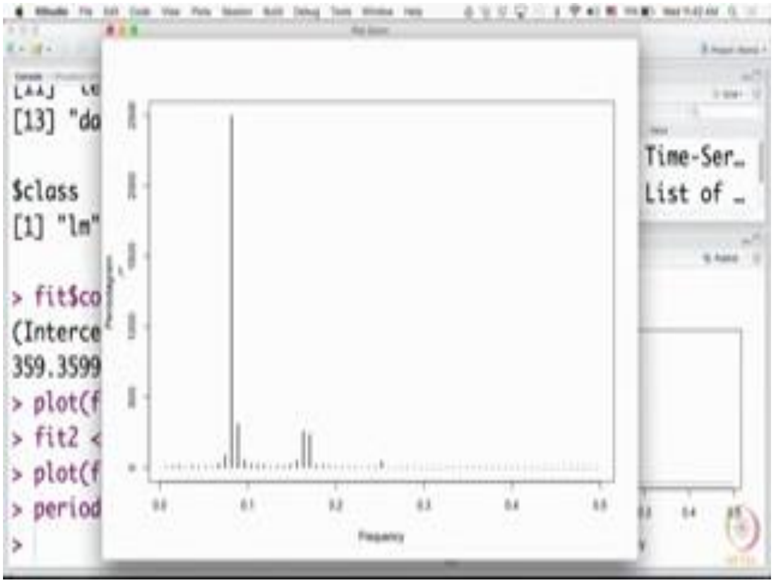
(Refer Slide Time: 15:05)



We can actually use the in the TSA package which we have already loaded, there is a routine called periodogram.

So, just to make yes we have loaded the TSA package, we can ask for example, with the first fit that we had where we only fit the trend remember we noticed in my two plots we notice for example, this periodic residual when we only fit the trend and now I want to know the periodicity. So, what I can do is I can generate the periodogram or and let me zoom out here to show you how it looks like.

(Refer Slide Time: 15:46)



So, what you see here is the power on the y axis and the frequency on the x axis right and you notice a peak do you notice a peak in the plot right there are other you know cousins to the peak, but they are kind of tiny subdued, but the predominant peak is somewhere at any guess.

Student: (Refer Time: 16:13).

0.8 roughly and that happens to be.

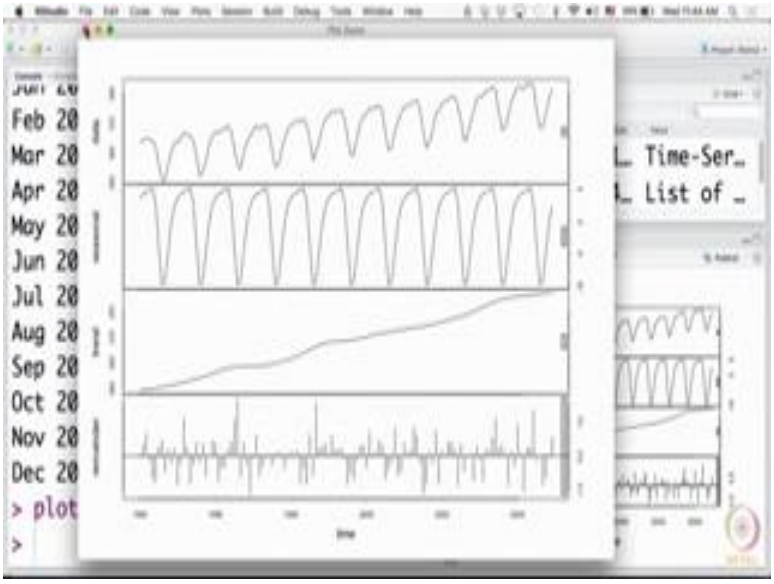
Student: (Refer Time: 16:16).

1 by 12 so, but confirms that there is a periodicity of 12 months or in in the sense one year, but in general they need not be the case and you may other harmonic such that which this we call as the for example, you can say 1over 12 the frequency at 1 over 12 is a fundamental frequency and then if you see there are 1, 2, 3 other this minor peaks, they may be the true periodic components present in the series they are not necessarily harmonics or they may be mathematically arte facts and we will talk about that later on what we mean by mathematical arte facts; at the moment I see a clear peak of 1 over frequency 1 over 12 period 12. And therefore you can go ahead with the second approach and now what you do is you have what we have done is we have fit a trend, we have fit the seasonal component and then we fit a ARMA model to the residual, if we

find that to be stationery. So, one has to go through the same procedure that we went through yesterday to build a model.

As I said there is another routine called STL; which does a beautiful job of decomposing this and it assumes that the periodicity again if you just say s dot window the seasonal window that is one of that is one of the arguments in STL, where you have to specify the periodicity or the length of the smoothing window for extracting the seasonal component when you do this I am sorry let us plot it.

(Refer Slide Time: 18:06)



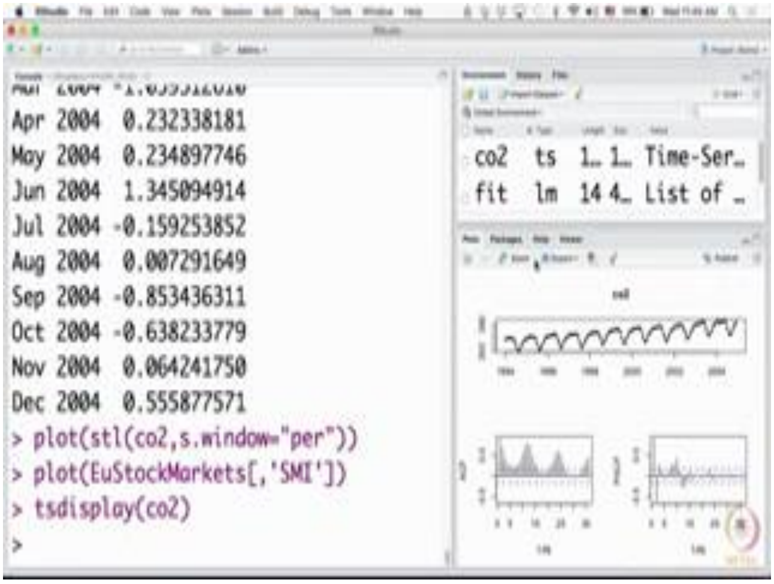
So, there is a very nice decomposition that it does, but it does not work all the time; however, it is a fantastic approach where it uses this non parametric idea that we have discussed earlier, it does not fit a linear trend necessarily, it does not assume any polynomial trend at all; it is extracting the seasonal component for you it is extracting the trend component of you for you and it is trend component is not obtained by parametric fitting you should remember that, it is obtained rather by the non parametric idea that I have talked about.

Of course it is not as straightforward as I have discussed; there is a more complicated method which we will not discuss in this course, but it is good to be aware of this method called STL and now you can access the individual components as you can see the series is on the top, the seasonal component is on the second panel, the trend is in third and the irregular component is in the bottom most panel.

So, this is one way of doing it, but provided you know the periodicity; you have to supply that information to STL. So, this is a general approach that we want to use, you fit if you if you want to follow a parametric approach you actually use this t s l m or l m and fit a trend whatever trend you feel is appropriate for example, you can try this on a sys stock market index example series that we have looked at before you know there is a trend and so on.

So, you can ask for example, here if you were to plot the sorry.

(Refer Slide Time: 20:03)



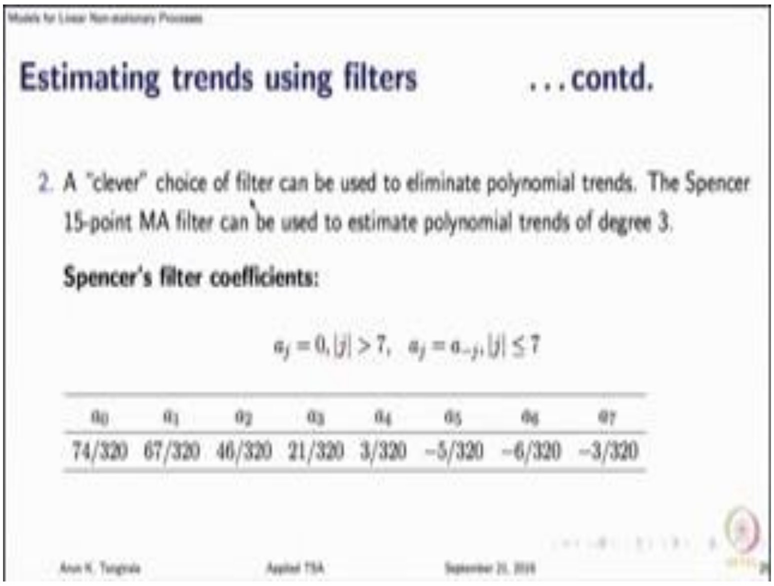
So, this is a series we have, here do we expect to linear trend to do the job no. So, then you have to fit a quadratic trend perhaps or maybe even a cubic polynomial, find it out what is the appropriate trend and see once we have removed the trend whether there is any seasonality and if there is none then; that means, you only have a trend plus this model.

So, this kind of concludes what I wanted to show as far as the trend analysis is concerned or handling trend type non stationarities is concerned, by the way I said the forecast package comes with a lot of other routine nice routines for example, there is something called t s display right and if we go back to our co2, it generates some it not only produces the time series plot, but also it gives you the ACF and PACF straightaway right the sample ACF and sample PACF are shown and what do you notice in the ACF? Some it is definitely there is a periodicity and also that the decay is slow.

Whenever you see such a phenomenon in the sample ACF, you should suspect some kind of a non a non stationery effect of the mean type, either a trend or an integrating effect right and one has to of course, distinguish between a trend type and an integrating effect because in terms of the nature itself they are different; one is the deterministic type, other is a stochastic type.

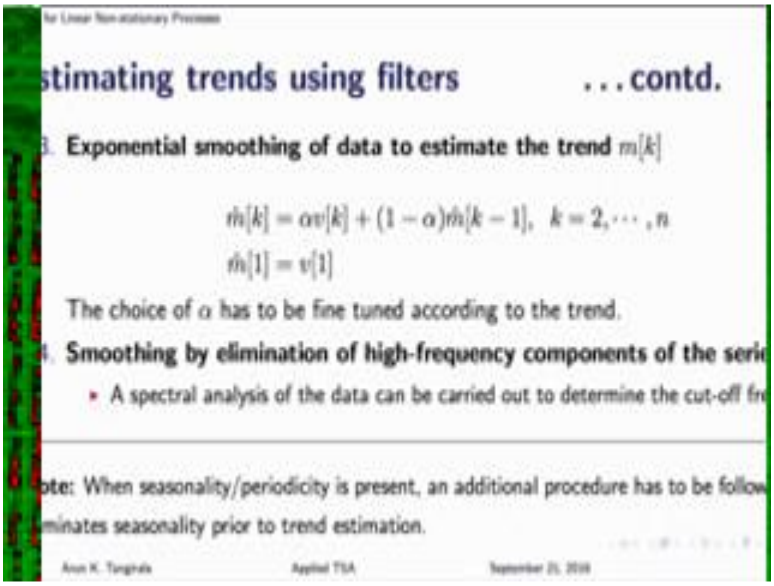
So, look through the some of the other routines in the forecast package and there are so many tutorials on the net, I strongly recommend that you go through the for example, quick r do not con get confused with quicker; quick r is a very nice website which has some very nice examples and demos.

(Refer Slide Time: 22:07)



So, let us get back to the theory part.

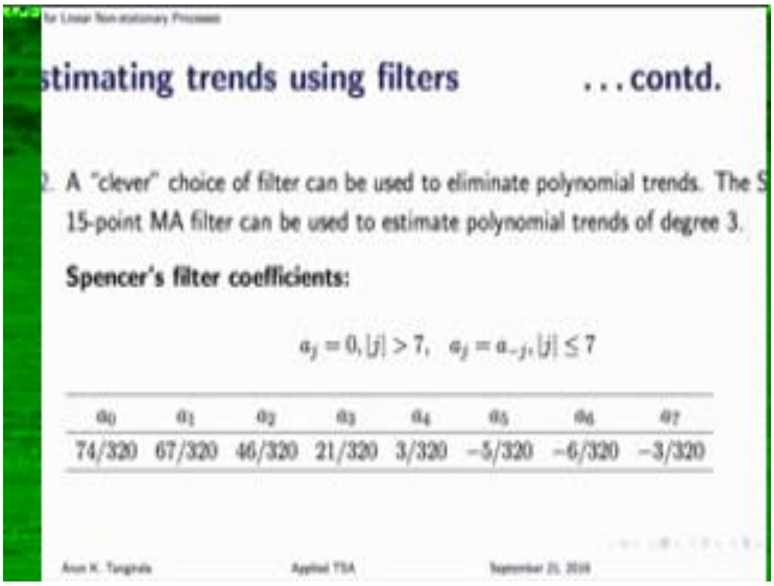
(Refer Slide Time: 22:10)



As I said you can follow a parametric approach or a non parametric approach and there are other approaches in a non parametric basket, one as I said is simple averaging, the second one which is used quite frequently is this so called exponentially smoothing and the way this m is generated is using this kind of a filter alpha times v k plus 1 minus alpha m k minus 1, and the initialization is done this way.

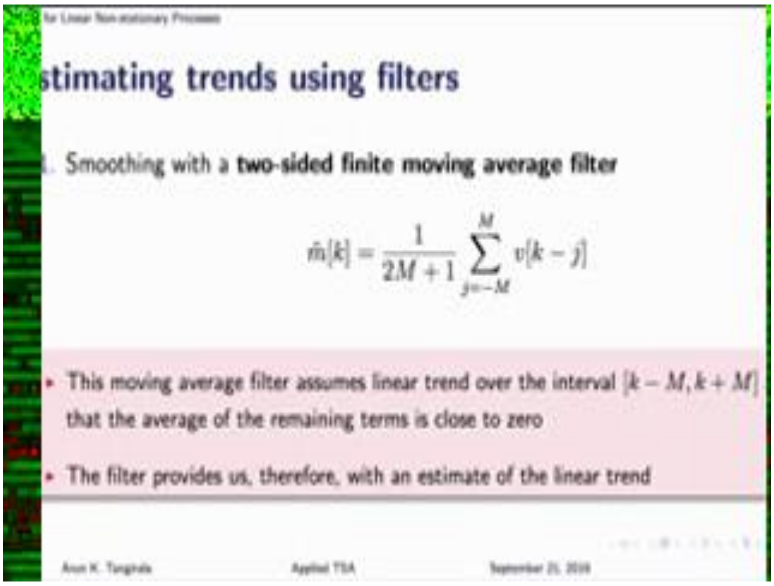
So, you start off with m of 1 being v of 1 means the first estimate is v itself; subsequently you apply this filter we will show later on that this is indeed a filter when we understand what is meant by filter, how to analyze the filtering characteristics of a filter and so on. So, we can come back to this exponentials smoothing equation and see what kind of a filter it is.

(Refer Slide Time: 23:10)



And.

(Refer Slide Time: 23:11)

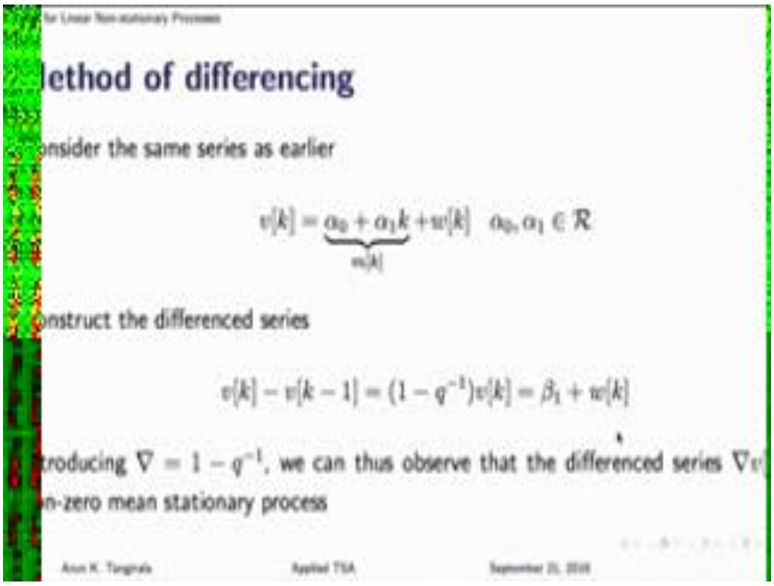


The other option is of course, to use some smoothing by elimination of high frequency components, essentially it amounts to applying some filter and you can apply there are so many filters available in the literature and you know if you come up with your own filter, you can name it after yourself also that option still exists.

But the bottom line is you want to carry out a spectral analysis of the data, to be able to get a feel of what whether there are periodicities, what kind of frequency components are present and so on and that will be the subject of our lectures for quite some time, but let us get done with this non stationarity business.

Now, the other approach of course, as I said is the method of differencing, which we also use in handling integrating effects, but in the method of differencing: the difference is that we do not necessarily estimate alpha naught and alpha 1, we just difference the series and then we find that.

(Refer Slide Time: 24:15)



The differenced series is stationery if the original one it as a linear trend, but with non zero mean; again here I have here beta 1 actually that should be nothing but your; what will be beta one in terms of alphas?

Student: (Refer Time: 24:347).

It is alpha one. So, it actually it should be alpha 1, but I have just kept it generic to show you that the differenced series as a non zero mean when the underlying when the given series as a linear trend; if it as a quadratic trend you have to difference twice, but there is one problem. In fact, this is this is wrong here there is it should be 1 minus q inverse w k; I will make that correction.

(Refer Slide Time: 25:00)

There is one problem with this difference in approach to handling linear trends. So, you have v k as alpha naught plus alpha 1 k, plus w k and what we are doing is we are differencing v k so that we have here alpha 1, but plus 1 minus q inverse w.

So, what is shown on the screen is not entirely correct. So, what we have managed to do is to get rid of the trend, but we have done something else, what is the side effect? What we wanted to do is we wanted to model v as a sum of these two components, in the earlier approach we fit alpha naught and alpha 1 and then eliminated it, then you would obtain an estimate of w; but here we have eliminated it by way of differencing and now we are left with not w k of course, there is alpha 1, but apart from that you are left with not w k, but 1 minus q inverse w k right. So, you will be fitting a model essentially to this series, what is that if w k is an ARMA process then 1minus q inverse w k.

Let us call this as w tilde this part alone as w tilde; w tilde unfortunately as a zero exactly at unity, is there a problem? We have assumed w k to be stationery and invertible what about w tilde? It is not invertible, we have forcefully introduced a zero and that is the problem with the filtering approach itself. Remember this differencing operation we have differenced here right is 1 minus q inverse, what we are doing is we are passing v k through an operation called differencing whose transfer function operator is 1 minus q inverse and then we get this.

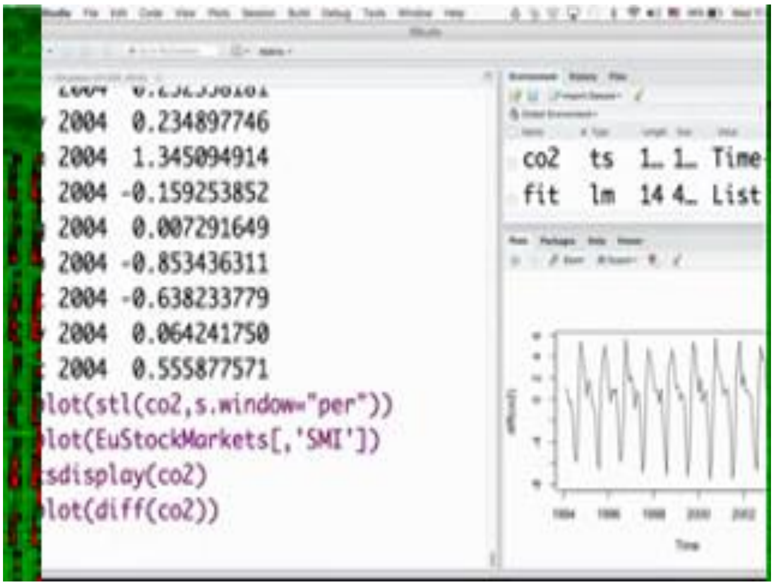
We will show later on that this differencing operation is also a filtering operation. In fact, I do not know some of you may must be familiar that this differencing operation is a high pass actually filtering operation, in this case in the earlier case we said we will eliminate the high frequencies to estimate the trend, but our goal is reverse here, we do not want to estimate the trend we want to get rid of the trend, trends are typically low frequency they are slowly changing with time.

So, therefore, what we are doing here by differencing v k is we are actually removing the trend, we are using a high pass filter, when I take any signal and I construct a differenced version of it, it amounts to passing the signal through a high pass filter; because when I take the difference of two points the rapid ones are highlighted, when I take the average of two successive points the rapid fluctuations are subdued; that is why averaging is a low pass filtering operation and differencing is a high pass filtering operation.

How do you know which is high pass and low pass? Well you just have to look at the net end result and ask what is it that I have retained and what is it that I have thrown away. In averaging I have thrown away or kind of subdued the rapid fluctuations and I have a smooth version, so the low frequencies are retained and therefore, it is a low pass filtering operation whereas, differencing is an operation where I am highlighting the changes when I look at the difference between two observation; that means, I am highlighting the fluctuations the rapid fluctuations therefore, I am allowing only the high frequencies to go through preference preferentially and that is why intuitively the differencing operation is a high pass filtering operation, we will show this theoretically later on that differencing operation is a high pass filtering operation.

So, this is completely in contrast to what we had discussed earlier; and unfortunately it is introducing a non invertible zero here right you cannot do anything about it therefore, when you have a trend deterministic kind of trend, you should try and avoid using the differencing operation to handle that, although differencing can mathematically get rid of the trend for you like for example, here I can go and plot that. So, diff routine in r constructs a difference series for you.

(Refer Slide Time: 30:05)



So, if you look at the difference series what has happened to the trend? It has vanished correct the only thing that it that it is left with is the seasonal component. So, we have managed to get rid of the trend, but underneath what we have done is we have introduced the non invertible zero and we do not want that correct, we want to work with stationary and invertible. Stationarity is not compromised, but inevitability is being compromised.

So, one has to be careful although the objective is to work with an stationery series how you obtain the stationery series makes a big difference. So, what we will do is when we come back tomorrow we will complete this discussion and talk about you know the methods of differencing in the context of random word processes; there differencing is absolutely the way to go, even if you know there is integrating effect, the differencing operation is one of the standard solutions, it does not introduce any spurious effects, but there also if you over difference.

If you have integrating effect of the first order, then one differencing will get rid of the integrating effect, but if you go further then you will again run into the same issue and that issue we call as over differencing. So, a differencing approach has to be used carefully you should know when and how much to use. So, we will complete the discussion tomorrow and then hopefully get started with Fourier analysis.