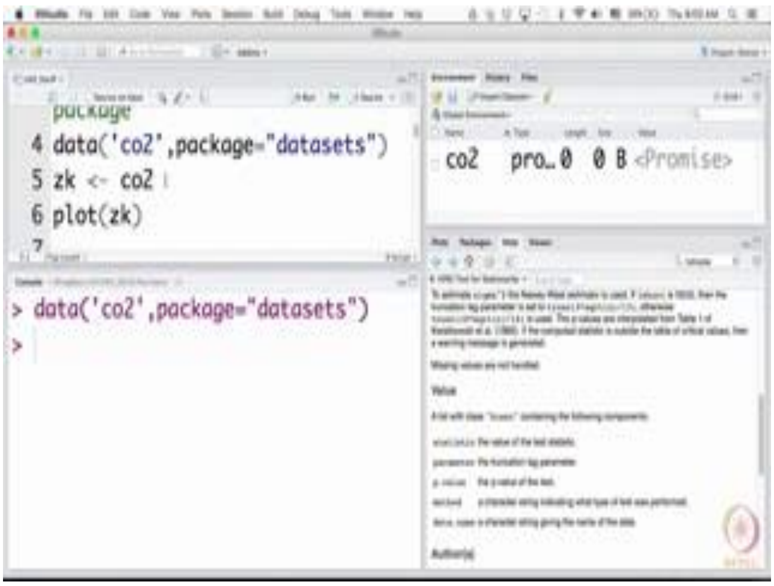
Applied Time-Series Analysis
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture – 55
Lecture 24A - Models for Linear Non-stationary Processes 3

Routine that we used from a forecast package yesterday, only allows you to extract simple trends whereas, you want to actually be able to extract or fit any kind of polynomial trends. And then we will we will make an important observation which will explain to you whatever at least an important point that we made yesterday which is that we can extract the trend by filtering or the other point being the trends are generally low frequency characteristics and that is why we either use a low pass filter to extract the trend or a high pass filter to eliminate the trend, a low pass filter will allow the low frequencies preferentially to go through whereas, a high pass filter will allow the you only the high frequencies to go through as a result of which the low frequencies are more or less eliminated.

Let us go through this example very quickly and then we will move on to the integrating type non statonarities and hopefully conclude with variance stabilizing transformations.

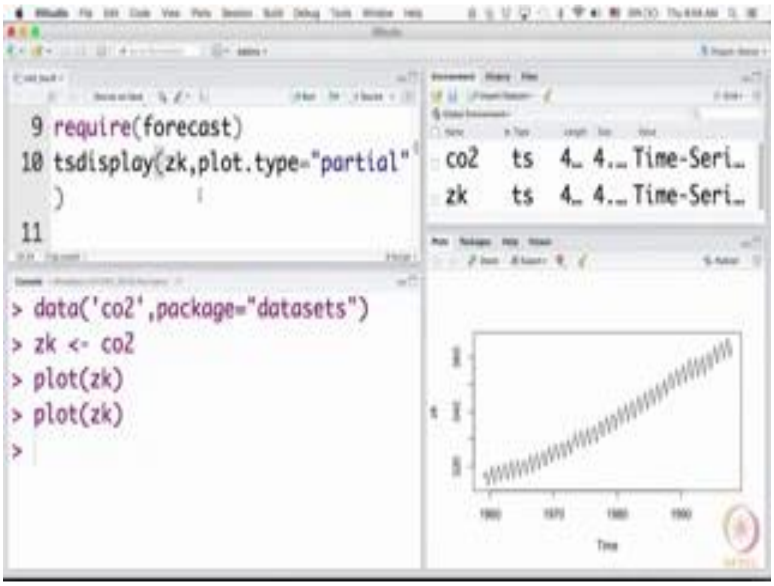
(Refer Slide Time: 01:24)



This is the co2 data, but this is now from the data sets package. So, yesterday we used the co2 data set from the forecast package, sorry the TSA package, but this co2 series also has a similar kind of features. In fact, let me actually go over this I am going to clear everything just to make sure that you start from a clean slate.

Let us actually run this particular line in the source code, I am going to post this script on the website, not to worry so, what this line has done is it has loaded the co2 data set from the data sets package that is the syntax for you to specify the package and then I am assigning this variable to z k just in case you happen to load co2 from another data set you do not want this variable to be overwritten.

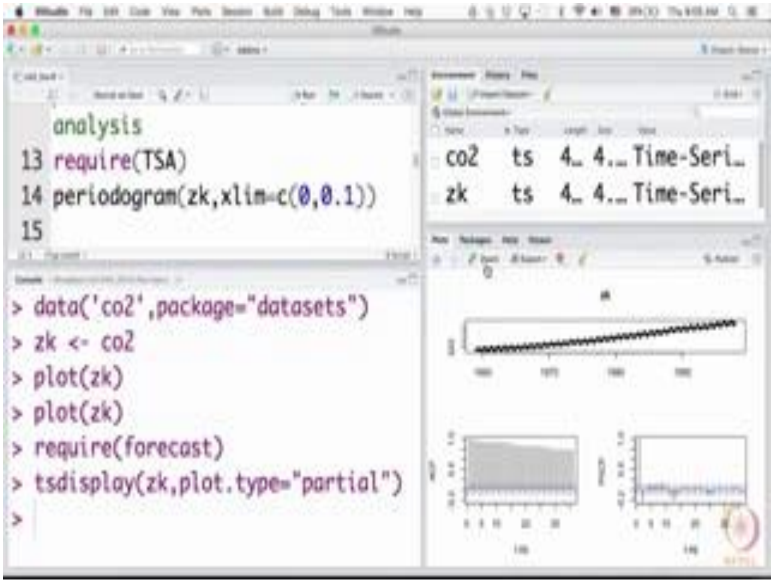
(Refer Slide Time: 02:36)



And now we can plot this, so sorry, I am going to close these variables. So, let us plot it again and there you go. So, you can see that this co2 almost has a same features, but do not get carried away in the sense that yesterday a linear trend allowed us to explain the deterministic characteristics, but we will have to wait and see if a linear trend will suffice on the face of it, what do you think, linear trend should be at least on the face of it, let us pretend maybe do not know anything and let us say that you have been, you have not listened to my cautionary note in that when you look at it freshly, it appears that a linear trend would be able to explain most of it, but that is only Maya that is what is called Maya.

Now what we will do is I am just showing you again the ts display routine that comes from the forecast package and I am just executing the script line by line so that you see what is happening.

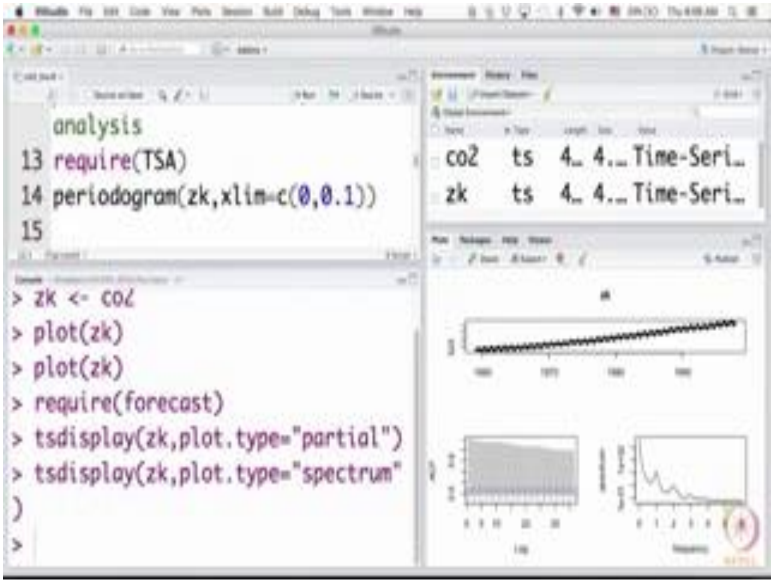
(Refer Slide Time: 03:39)



You have seen this before, sorry for the other data set and once again the Acf shows a slow decay, but in practice just looking at the ACF and I am looking at the slow decay of it without looking, sorry, without looking at the series, we will probably compel you to believe that there is an integrating effect because we have talked about this when you have integrating effects, one of the ways to see if there is an integrating effect is to look at the ACF, but the series clearly says there is a trend. So, you can rule out the integrating effect for now.

Now, there could be integrating effect on top of a drip. So, again do not get carried away, but definitely what we are sure from the series is that there is a deterministic kind of trend and the p ACF if giving you some information I should also tell you that this ts display allows you to display both the ACF and the spectrum.

(Refer Slide Time: 04:53)



For example, here in the ts display, I have given type as partial, but instead I changed this to spectrum and then execute it, you can see that it shows the series the ACF on the lower left corner and then the spectrum as I have explained yesterday is a plot of the spectral plot is a plot of power versus frequency.

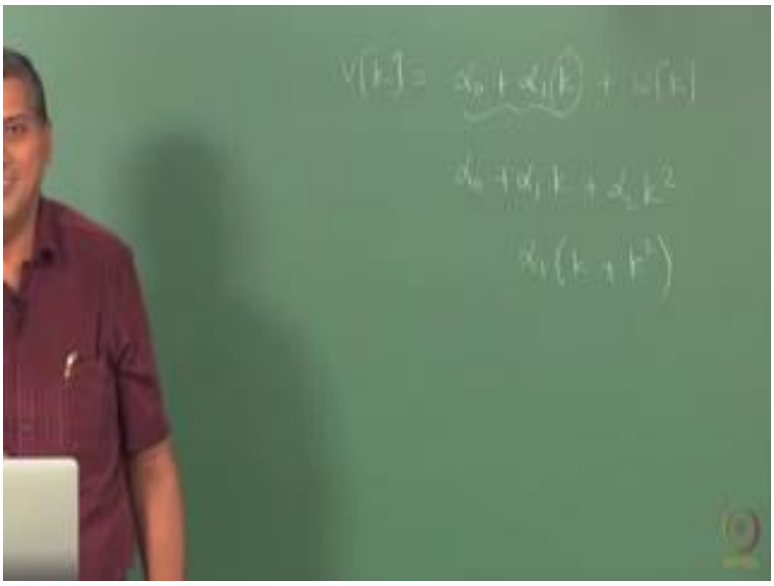
What do you see predominantly here? Even you look at the spectral plot, what can you infer about the series you think? So, there are 2 peaks, correct which means there could be a fundamental and harmonic, we do not know, but what else do you see? So, if you were to look at if you have to ask yourself, what is the you know contribution of frequencies which regime contributes predominantly to the overall power of the signal, we have not defined formally what is power we will do. So, low frequencies, this is because the trend is so dominant and I will show you once we remove the trend these low frequency contributions vanish.

So, that is the and sometimes the trend can be so dominant that it can shadow the periodic component here, of course, the periodic component is good enough to manifest as peaks in the spectral plot, but sometimes the trend can be so strong that it can overshadow and therefore, you should be careful when you do spectral analysis, you want to remove the trend before you do an if you before you carry out an analysis for periodicity.

These are some simple examples which convey very important aspects of time series analysis. So, we will keep this in mind whatever we have seen visually is being confirmed very nicely by this spectral plot and to a certain extent by the ACF plot, very good, any questions on this from the other room? So, now, let us move on and we will skip the periodogram part, you can execute it, it is going to show more or less what we have seen in the spectral plot.

Now, let us actually construct the time vector remember now we want to care construct linear fits or maybe quadratic polynomial fits and so on. So, you need this vector of time instance in the linear trend case we postulate this model for the trend plus of course, a stationary process.

(Refer Slide Time: 07:22)



I need this k vector so as to carry out linear regression and that is what we are doing here. So, the tvec or the kvec, whatever you want to call, NROW is the built in command in r which will give you the length of z k, do not use the lower case, the lower case assumes certain type of objects, but this upper case NROW works for most of the objects. So, NROW of z k would be the length of the series we want from one to NROW, you can start from 0 to n minus 1 as well, and it does not matter.

Then the intersect term would change, that is all. So, we will do that and then what we do is we fit polynomials of different degrees. So, here this is the first fit, linear fit, oops something is wrong, sorry, there is a mistake, here should be z k that is why it is always good to begin from a clean slate because I may have a variable v k and it will do all the things for you and then half an hour is gone paper is lengthy. So, this is our model with a linear fit and then the quadratic fit observe how I am constructing the quadratic fit I am using this I of tvec plus I of tvec square now that tells l m that there are 2 regressors k and k square if you were not to do that, but is if you were to give tvec and tvec square it is just fine I am happy with that, but it will cons consider k plus k square as a single regressor.

You may want to fit, you may mean to fit this kind of a model, but if you use the wrong syntax then it may fit a model of this form. So, using I we are telling l m that each of this is a separate regressor and of course, intercept term is assumed by default if you do not want the intercept term then you use the minus one in addition in the formula.

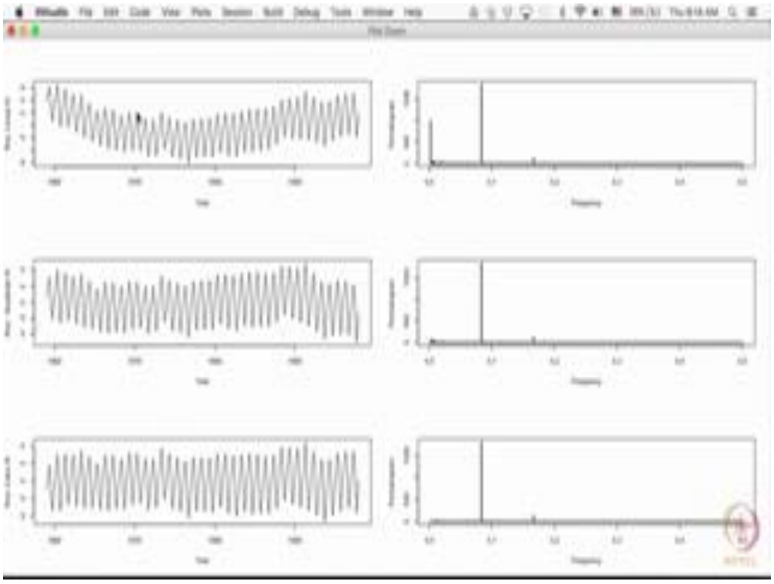
And then we do the cubic fit, so now, what we want to know is which fit is good, whether the linear fit is good or a quadratic or a cubic and so on, one of the things that you should make a habit is as soon as you fit you want to for example, pull up the summary for the model that you have fit. So, a summary of tr fit for example, for linear trend gives you these results here it shows the coefficients we have talked about it before these are the estimates here of the intercept and this is the estimate of the alpha one and on the sidelines you have these standard errors.

Of course, as I told these standard errors should ideally be used after you are convinced everything is explained nicely, but we will slightly deviate from that and still peak into the error you know just a curious mind we want to see how much is the error whether the error is relatively small and it turns out to be relatively small and even if you are in doubt let us say you have forgotten everything you know the stars right so; that means, these are the significant one these both parameters have to be present in your model can be present in your model.

And then of course, now there are other things that are thrown out for you multiple r squared then adjusted r squared we will learn those when we learn the least squares methods. So, likewise you can actually pull up the summary for the quadratic fit as well and we want to know if inclusion of the quadratic term as actually made any difference or not in the sense has is alpha 2 now a significant parameter or not in fact, alpha 2 had.

It turns out that they are right, but this is one way of looking at it as I said you should not just go by the errors there are some assumptions in calculating these errors that may not hold good for the series it is best to plot and see visually for yourself what is happening. So, let us plot the residuals and see what happens. So, what I am doing is I am going to split this screen into 3 by 2 matrices graphic screen and then I am going to plot the residuals and the period spectrum of the residuals right if the fit is good then what I should be seeing visually in the residuals is only a sinusoidal or a periodic kind of series plus of course, some randomness. So, let us execute this.

(Refer Slide Time: 12:30)



Now, I am going to just source the entire file and I am now going to zoom out here so that you can see what is happening. So, here is the plot the top one is the residual coming out of your linear fit the top left and on the top right you have the periodogram. So, what do you see with respect to the residuals from the linear fit? Do you still do you see that there is a trend? There is a low frequency kind of thing mean you can attribute that to a low frequency sinusoid if you like or you can say no it may be there may be a quadratic thing it is actually very slow and you can see that here in the periodogram plot the it is quite different from what you saw earlier.

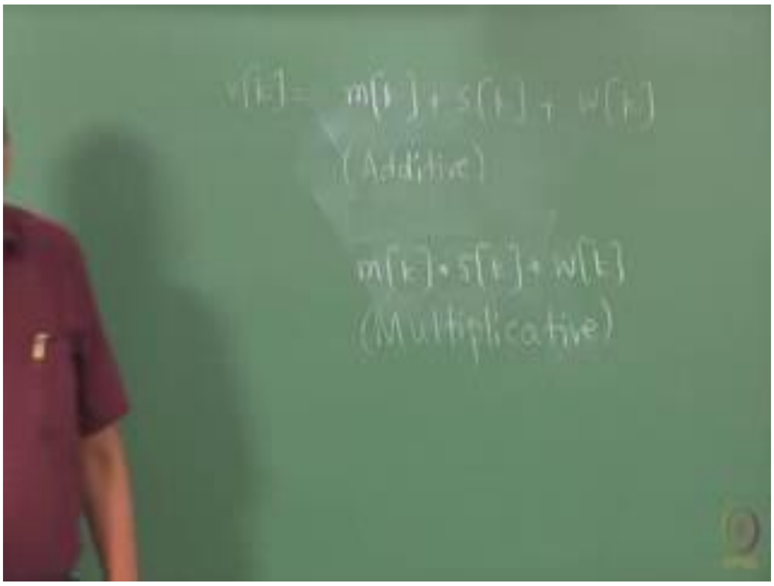
Now, the periodic thing is showing up, it is head, it is like now I am the dada the other guy is gone, but there is still this low frequency component present which is shown by the peak, there at the low frequencies and that peak could be either due to a very low frequency sinusoid or some trend quadratic trend whether which one is correct is what we want to check and that is why we go through a quadratic fit right once we have a quadratic fit the residuals are shown here and still you know there seem to some very low frequency component present this could be due to a cubic one or maybe an extremely low frequency sinusoidal and you can see this very very tiny Eskimo fellow here.

Now, you all almost even maybe very very short here, but predominantly now you are left with the periodic component and we go further to the cubic one and we have more or less reached the stationary not exactly stationary, but the stationary plus sinusoidal and you can see now in the spectrum plot periodogram plot that there is this fundamental frequency of one over twelve that is the cyclicity of twelve followed by an by a harmonic these are the 2 peaks that we saw earlier.

This is how you systematically figure out what trend is suited for the series what is the seasonality and. So, on as I said there are 2 routines called decomposed and STL there are 2 routines in r the decomposed does this for you, but in a non parametric way that is and so does STL, we have adopted a parametric route here and we have not assumed seasonality we have relied on the periodogram to tell us what the seasonality or the periodicity is I keep using the term seasonality, but remember that seasonality is much more than periodicity whereas, many of these methods by default assume that you have to supply this periodic this period periodicity information.

Now, with the decomposed routine for example, if I have were to use decompose on this let us see what happens. So, we specify now in the decompose we can specify whether we think of the series as an additive model we know what is an additive model right additive model is where you assume the trend and seasonality and the random component add up.

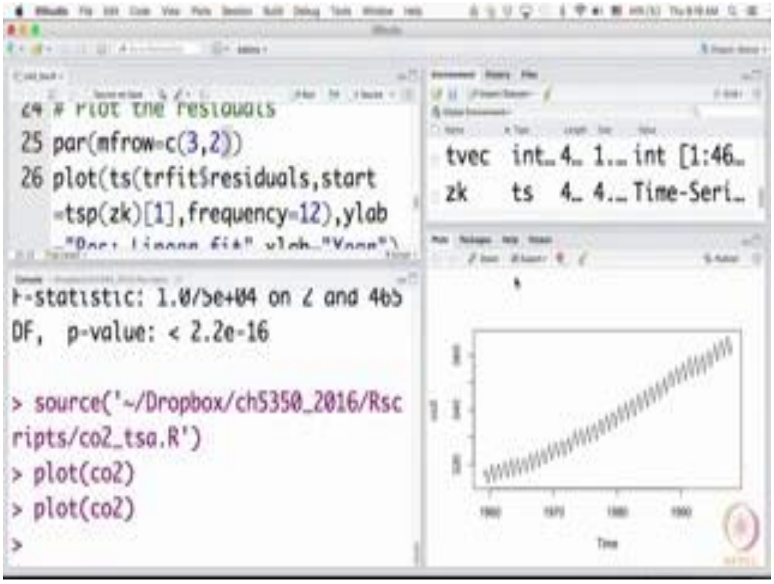
(Refer Slide Time: 15:51)



M k is the trend and then you have s k the seasonal component and then the stationary component this is your additive model whereas, in the multiplicative model you assume that the series is a product of these 3 components.

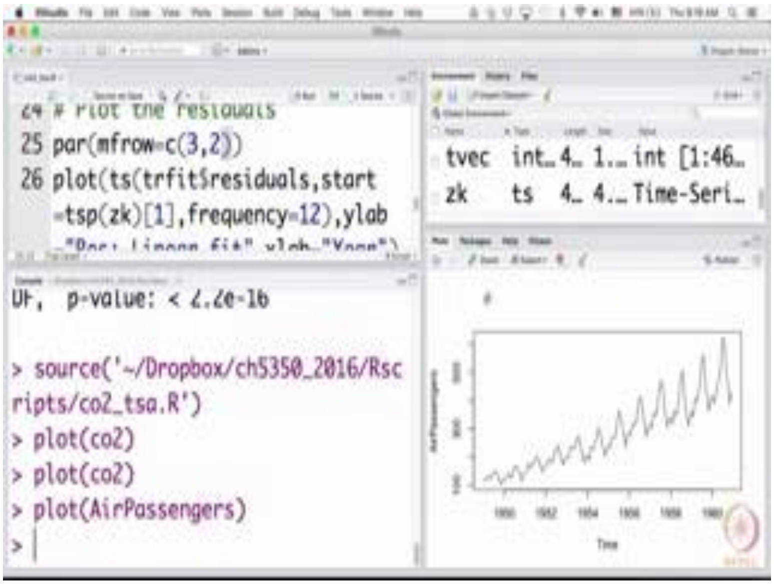
Now, you can specify, how do I figure out whether the series whether the additive model or the multiplicative model is suited, which one is better suited one of the ways to figure out although it is not a very foolproof way of doing it what this model says. So, let us look at this model and then I will show you the example this model says effectively is that the amplitude of the seasonal component if you were to think of this only as a seasonal component the amplitude is a function of time and also has some irregularity whereas, this model says all the 3 are separate on their own the trend does not affect the seasonality and you know one of each one is really does not affect the other so that is the thing.

(Refer Slide Time: 17:24)



If you look at the series that we have been analyzing sorry I will come back to the decomposed part let me close this where is this green as blue came. So, if you look at this series here what do you think which model do you think is suited the additive model or the multiplicative model right correct. So, the additive model is suited why because the amplitude remains the same for the periodic part.

(Refer Slide Time: 17:49)



On the other hand, if you were to use the multi look at the air passenger’s series, what do you think here a multiplicative model is better suited?

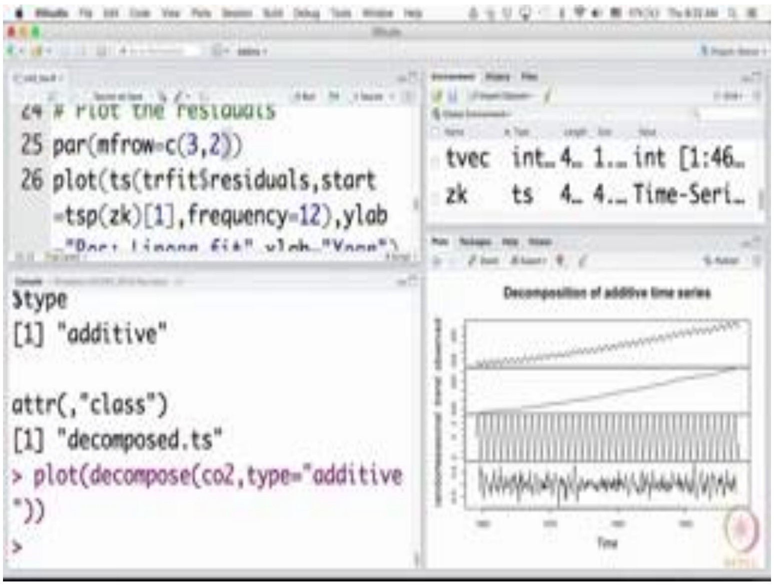
Lot of times you can, but sometimes you may have it does not mean that if you use a multiplicate additive model here you are committing a crime and the police will be at your door nothing like that you can use it, but then see modeling is all about explaining the data whether one person has a better way of explaining than the other whether you have a bet a model that mathematically explains the process and more consistent with the physics also if you know something about the physics we say that the multiplicative model is better suited based on the features that we have examined.

Look at the power of the visual inspection our brain is able to do an amazing amount of signal processing just in a, you know in a flash of a second. So, we feel that the multiplicative model is better suited what happens if I fit an additive model to this do you think something will happen, what could happen? Somewhere these 3 components have to explain the series right in the multiplicative model, what happens is you can expect w to be stationary and m k to be having a cert I mean you can see the trend and then of course, there is this periodic component if I were to fit an additive model somewhere these features have to be absorbed.

The seasonal part is on, it is own the trend is on it is own, I can fit a trend to this and then I can fit a seasonal component to it, but the seasonal component I can see the periodic component I can see is actually as a changing amplitude the way it is depends on how I look at it how do I explain that because in the additive model the amplitude is kind of fixed where do the changes go and sit in part goes and sits in the trend and part of it will go and sit in w. So, the w and m that you get can be quite different the periodicity remains the same in both.

Let us see what happens if I fit an additive model and a multiplicative model to the series here and then we will move on to integrating effects. So, we use a decompose and I should tell you that STL is a more sophisticated routine, but explaining the algorithm behind STL is beyond the scope of this course I will not go into it again the in more or less the inputs are the same whatever you give to decompose more or less have to be given to STL, but the algorithms are different and in decompose you have the facility of saying whether it is additive or multiplicative whereas, in STL I did not find any option.

(Refer Slide Time: 20:54)



Let us try for the co2 for example, and specify clearly that you want an additive model and hope sorry we want to plot it. So, here is the plot of the decomposition, but this is done in a non parametric way like the way I explained yesterday how is the trend obtained? For example, in this kind of an algorithm the trend is obtained by first knowing the periodicity. So, it assumes that you know the periodicity and that is actually in this case where did I specify the periodicity in the routine well there is the co2 that I have supplied is a time series object there a time series object has certain attributes start time end time and how frequently we are obtained and so on.

If you look at the frequency for example, let us look at the TSP right here, the last one is a frequency. So, it says 12 and it assumes that to be the periodicity. So, given the periodicity it extracts the trend how does it do it we know that a periodic component has a 0 average assuming it to be 0 average over one period it constructs estimates of m k by using a window of length of that period and then it moves on you understand yesterday we talked about this idea of constructing local averages to estimate m and that is exactly what it is doing, but here when you have a seasonal component you cannot use any window.

If you use a window different from the period then that will affect the estimate of mk, we want to use a window such that locally the average of s k is 0 and more or less the average of w we have assumed w to be 0 mean. So, locally hopefully that will actually produce a fairly low estimate of low component of average of w. So, predominantly what you will get is m k the difference between decompose and STL is STL does this in an iterative way one of the main features. So, once you have the trend component you can also given that the periodicity it extracts the periodic components subsequently and whatever is remainder is w k very straight forward.

Now, if I were to do a multiplicative fit to this. So, let us actually quickly ask what decompose would do if I were to postulate a multiplicative kind of model now is there any difference per say that you notice. So, let me actually go back and here is the only difference of course, is in the title I do not want you to tell me that that you what is the difference that you predominantly notice in the w case and what else you amplitude of the correct of the random the irregular component and perhaps the seasonal 1 2 because it is a completely different model, but features wise it has extracted the trend and the periodicity you may notice a big difference with air passengers I am not going to go over it, it is a simple exercise go back and try it out just out of curiosity.

Now, the generalization of this multiplicative model are your serima models seasonal arima models which can actually take into account integrating effects seasonality and so on, this multiplicative models were proposed by box in Jenkins in early seventies and then since then they have become very popular. So, you know now at least how to handle different non stationairites, at least these classes of non stationarities where you have trend and periodicity riding on top of a regular stationary series, any questions.

Student: Sir is there any (Refer Time: 24:48).

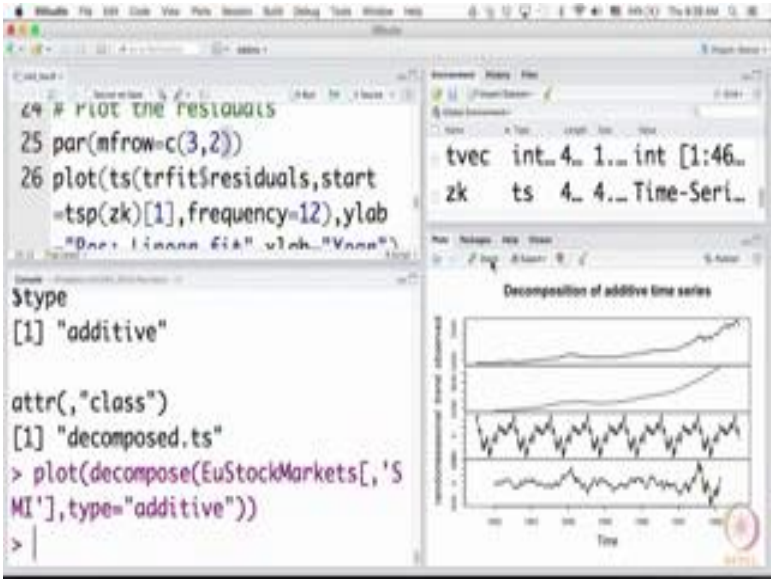
No, no, when you ask such questions you have to postulate really what is it that you are testing, so, you are asking whether I can determine if there is a trend type non stationarity? Yes.

Student: (Refer time: 25:15).

Then you have to conduct, you can conduct, now further tests of non stationarity on the regular component, but then you have to be aware of what different types of non stationarities are there what is that non stationarity that you are testing for unfortunately the tests have to assume something non stationarity they are devised for a particular type of non stationarity for example, I can ask now if there is a random walk kind of phenomenon in there and then there are unit route test random walk test for that or I can ask for example, if the variance is changing with time right and that can also happen for example, if maybe if I use SMI, sorry, e u.

Let us ask what the SMI does decompose does to SMI? So, let us use for example, an additive model in this case. So, look at this series here, it has grabbed the seasonality information from the object time series object and it has got you the trend again in a non parametric way the seasonality in a non parametric way and then wk.

(Refer Slide Time: 26:18)



Do you think wk is stationary here? What do you think? Something is wrong with it in with respect to stationarity, maybe it is variance is blowing up mean does not seem to be blowing up.

Maybe there is a variance type non stationarity and which is what is considered in the world of heteroskedasticity, you may then you have tests for heteroskedasticity also available, but it is a good question, simply assuming the wk to be stationary would be kind of misleading and probably erroneous. So, one has to conduct some tests of non stationarity on wk as well.

Student: (Refer time: 27:20) we are assuming it as a 0 mean.

Which 1, wk?

Student: wk, if it is non zero mean (Refer Time: 27:27).

No, if it is non zero mean it would be absorbed in your m k the intercept term of m, k that is not an issue, the more serious issue of the integrating effects or the heteroskedastic effects and so on, any other question from the other hall, yes.

Student: now we take the random data (Refer Time: 27:49) periodicity if we use non parametric method.

No, you would not, that is why you have to first know it upfront that is the disadvantage of the non parametric method.

Student: (Refer Time: 28:00).

No. So, you have to do some preliminary analysis come back and you know use that that is one of the demerits of using this methods you have to know, you have to have some information a priori about the periodicity otherwise, it will not be able to extract or you can say, I will use this window for averaging, for getting the trend that is an indication of saying, there is a periodicity of that much, but in the series that we have looked at a kind of new, but that is why I showed you the periodogram analysis. You remove the trend you for example, perform a parametric analysis get an estimate of the seasonality then go back perhaps use decompose. And that is why the spectral analysis is going to come up very soon to help you detect the periodicity in some sort; any other question from the other hall.