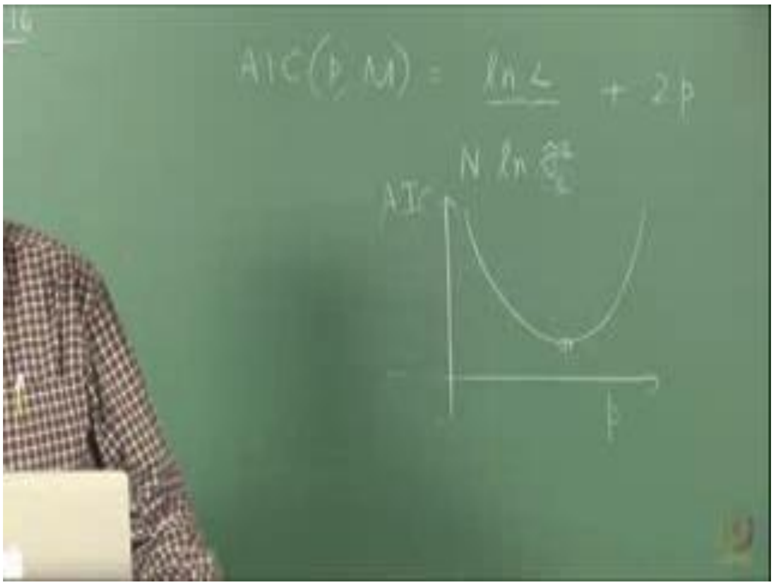
Applied Time-Series Analysis
Prof. Arun. K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture – 53
Lecture23 A - Models for Linear Non-stationary Processes 1

Very good morning, yesterday we went through a session where I showed you how you can systematically develop a time series model by following certain procedure and towards the end there was a question from one of students; after the end of the class as to what this A I C that we had use Akaike Information Criterion is actually quantifying.

(Refer Slide Time: 00:43)



So, just for the benefit of everyone although we will not discuss in detail, I just want to mention that this A I C is actually a function of the number of parameters P and the model also. It is a function of both your P and the model M and there is a mathematical expression here for computing A I C; depends on how you look at it for example, one of the expressions that is originally given by Akaikeinvolves the likelihood function which we have not discussed yet but we will soon do so.

So, the first term is log likelihood, if you do not understand what is log likelihood then you can think of this being replaced with n log sigma square. So, let me put it this way sigma square e hat alright, the sigma square e hat is a estimate of the noise variance that you get plus another function; another term which contains a number of parameters in the model. Rather than worrying about exactly the exactness of the expressions given there, it is important to interpret each of this term the first term signifies the fit that the model has achieved that is the goodness of fit. If the model has fit the data very well, the sigma square e although we say here sigma square e had truly what happens is there it is nothing, but the average sum square prediction error. So, 1 over n sum square error over the data points of your fit; when I say error here prediction error.

So, as I increase the model order for a given type of model, the first term actually goes down, but then you are increasing the second term because you are increasing the number of parameters in your model and what does the second term signify as we learn later on, it signifies the amount of error that you are incurring in the parameter estimates; typically as the number of parameters that you include in your modeling increases, the error in a parameter estimates will also increase. So, the second term is a quantification of that and now you can see there is a trade of between this term and the second term here.

As you increase the model order, the first term lowers but the second term starts to shoot up and you are looking for that point; that troughed; if you are to plot A I C versus P that is a number of parameters; ideally Akaike aspects to see this kind of point there and we are seeking this point because this is the kind of optimal trade of that you have achieved, if I have a model with very few parameters or kind of no parameters, then the first term is going to be very high because you have not fit any thing or may not done a proper job of fitting the data, but as you walk on this axis; walked down, you are increasing the number of parameters there by having more and more parameter power to explain the data.

Therefore, the first term starts to decrease and at some point that decrease in the first term becomes marginal; that means, let us say in the yesterday's example we fit a fifth order of a r; we could have gone to a sixth order a r and achieved a lower prediction error, but then the improvement that you get in that respect is going to be insignificant compared to the shoot up in the errors that you are going to see, as you move from the fifth to sixth order.

So, that is the trade of that you are increase, so the extreme cases where you have large number of parameters in the model; heavily over parameterized model which does a very good job of fitting the data, but what happens is the error in the parameter estimates is so high as a result the A I C start to shoot up; it takes a u turn. So, you are looking for this point of u turn and these all idealism in practice you may not see such a sharp turning point; you would rather see a flat region; kind of semi flat region and then you have to pick based on (Refer Time: 05:26) and so on.

So, that is the principle behind A I C, but we will talk more about it later on and this likelihood function will become clear once we discuss a concept of maximum likelihood itself. So, let us get back to the discussion and there is one more point that I want say that we have use certain tools yesterday in building the model and as I said there are some minor issues with using this tools which can go unnoticed by an ordinary user. So, I would strongly recommend that you visit staffers website, where he explicitly talks about these issues concerning the use of ACF, PACF and even the (Refer Time: 06:14) have be used that routine and a few other routines that you would use in future. So, it is important to be aware of those issues and staffer has some remedies in terms of new routines and so on.

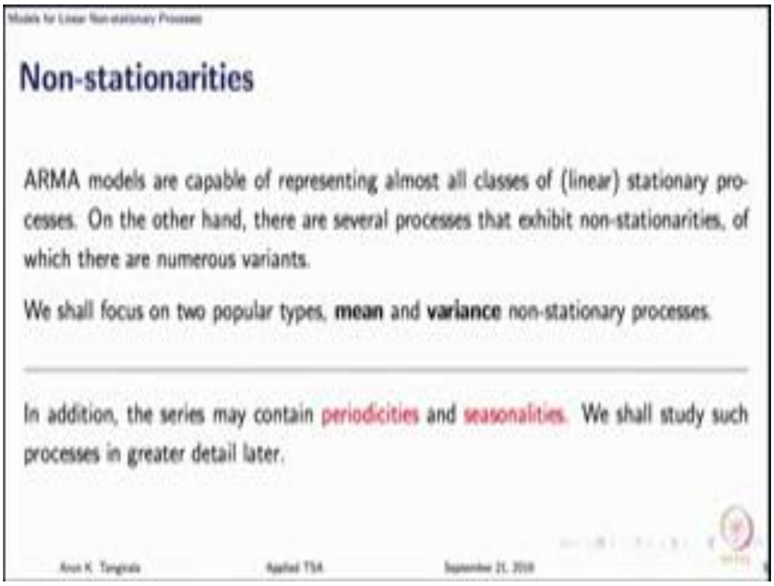
But please go through that page, you just go to staffers website and look for issues with r and of course, he has some very nice sectarians sarcasm where it is really fun to read the way he brings up the issues and so on. One of the things that you also point sort as we know when you load new packages in r, there are going to be routines whose names overlap with those in the existing packages and therefore, when you load a packages it says these are being massed and so on and his take on that is one day r will actually get into an infinite loop, where there will be some routine which will mask and then again another routine which will mask and gets into infinite loop and that is how r will end its life something happen.

So, anyway there is another package that I am going to talk about today which is very useful and you should install it, but let us start our theory session and then in the course of the lecture when I demonstrate how to fit for example, trends or how to identified seasonality and so on; at that time I am talk about that packages.

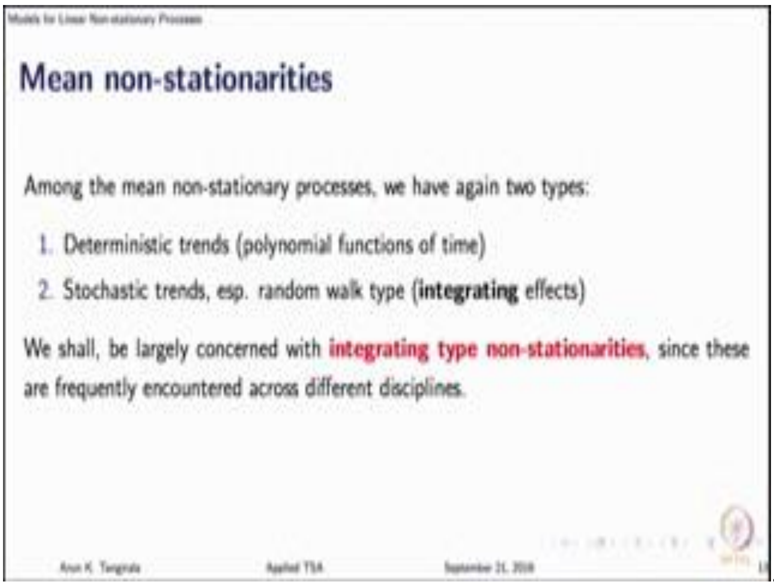
So let us get started with the theory part again; until now we have assumed stationarity, but as we have discussed before many processes do exhibit non stationarities and that is the topic of today's lecture and there are different types of non stationarities; just like there are different kinds of non-linearities. So, you can have mean non stationarity, you can have variance non stationarity and within mean non stationarity you can have two different types; deterministic one and stochastic one and then to top all of this, you can have periodicities or in general seasonality; seasonality is a generalization of the concept of periodicity.

So, what will do is we will actually start with the mean type non stationarity and discuss the deterministic one first.

(Refer Slide Time: 08:25)



(Refer Slide Time: 08:29)



So let us do that; when you look at the deterministic trends typically you are looking at polynomial functions of time, whereas when you are looking at stochastic trends, you are looking at typically random walk type of effect and we have talked about this before. Now what we shall do is primarily focus on integrating type effects, but I will also talk about how to get rid of trends, there are many series which is exhibit trends and there are other ways of handling deterministic trends which I am going to briefly talk about, but the focus is going to be on integrating type non stationarity, in other words we will mostly look at ARIMA models and the advantage with working with this ARIMA models is that to a certain extent they can also handled trend type non stationarity although this not recommended to use that I am going to talk about certain settle aspects of working with ARIMA models.

So, let us move on and before we do that also sorry there is one another point that I want to make, when we talk of deterministic effects there are again two types; one as we have already noted down polynomial functions of time, those are the kinds of trends we have seen this before. Then you can have another kind of deterministic effects, which are due to exogenous effects that is some external variables is introducing these kind of deterministic effects and in such cases, you will have the knowledge of the external variable and therefore, you will include that or factor that in your model as well.

So, in other words no longer do you consider we as purely a function of time or its past, but also now as the function of this external variable and that is essentially the subject of system identification that is how time series community view system identification as. It says that whatever inputs you know are producing changes in the response variable that you have, can also be factored for example, in ARIMA routine that we used yesterday, there is an option to include exogenous variables as well and when you do that; what the models that are born are called ARIMAX models; the X standing for exogenous and that is why lot of normal (Refer Time: 10:56) system identification also is born from the time series literature.

So, you have ARX models for example, in system identification ARX models; it is an A R model with exogenous effects or you have R max models which is an ARMA model for the noise part and then a exogenous effect being included in the model. The exogenous effect is also typically considered as being model as l t i system; that is the input or the external variable that you have is assumed to be effecting your response through an l t i system and that is the subject of at least linear system identification, but we will not go through that.

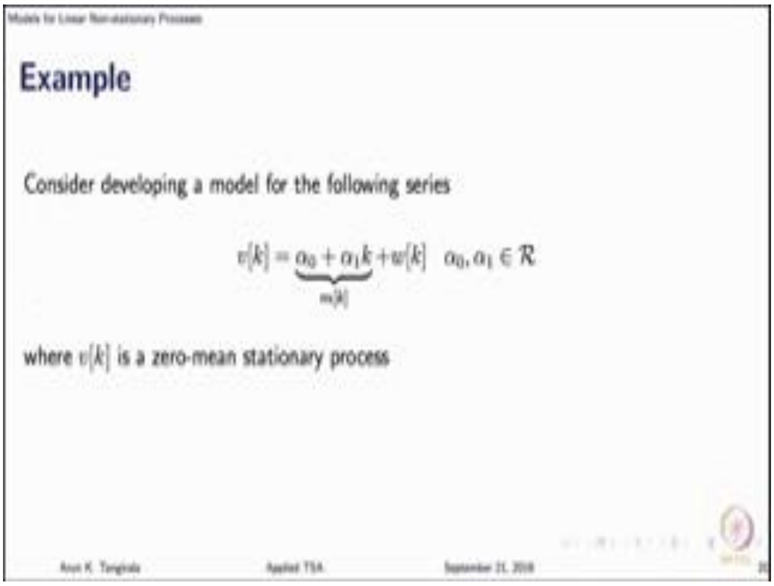
Never the less; if you are given exogenous variables, you can still use that information and factor it in your ARIMA routine and your ARIMA modeling essentially.

So, let us come back now and talk about time trends that are purely function of time; we do not worry about exogenous variables now.

(Refer Slide Time: 12:00)

And there are two approaches to handling trends; one is a of course; fitting a polynomial to the given series and then working with the residual. The other approach is eliminating the trend by the use of a suitable filter and we will discuss both these approaches now.

(Refer Slide Time: 12:23)

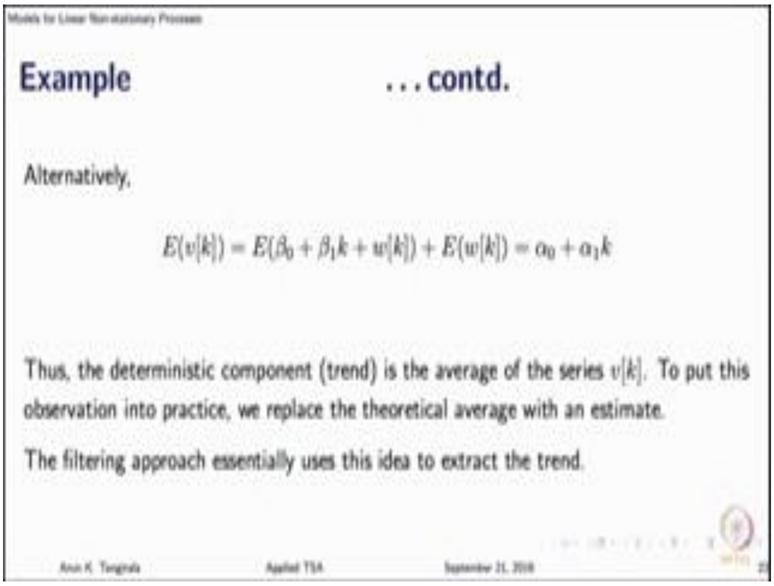


So, for example, let us say there is a linear trend; at v k sum alpha naught plus alpha 1 k plus some w k and wk can be modeled as an ARMA process; invertible stationary invertible ARMA process and we can assume it would be 0 mean without loss of generality.

So, the procedure here or the practice here is to fit a linear model in time to this series that is given to you and then work with the residual; how do you fit this alpha naught plus alpha 1 k using standard leak square methods. Of course, there are satellites because when you are using leak square methods, the optimality the efficiency and the consistency of leaks square methods depends a lot on the nature of wk, but we ignore all of that; we will assume that you will get optimal and efficient estimates and so on. Typically there not violated at least in this case so much and just fit a linear regression kind of model and that is achieved by l m; in r I have already shown to you a routine called l M which does the linear regression for you and there is a better routine that I will talk about very soon in another package, which make your life bit easier in terms of fitting this trends.

So what you do; once you fit a straight line here then you work with a residuals, how do I know it is a linear fit right; it could be quadratic trend, it could be some other polynomial trend; you will have to make a visual inspection guess and then relook at the residual if you still find the trend then go back and make corrections to your polynomial trend and so on until you feel that the series is stationary. There are test again for this which will tell you whether the underline, the series has the linear trend or not and so on, we do not go through such test here, but you are most welcome to look up the literature.

(Refer Slide Time: 14:39)



Now, alternatively as I said you can use the filtering approach in r; I will demonstrate how to do the former one which is the fitting the trend shortly, but let us quickly discuss this second approach as well. So, if you look at the second approach; I have beta naught and beta 1 that should be actually alpha naught and alpha 1. so just I will make that correction later on. So, when you look at the expectation of v k, you can see that it is simply alpha naught plus alpha 1 k.

Now; that means, in some sense the deterministic component is the average of the given process correct; this is the basic idea in using the filtering approach. What is done in practices is you take sample mean of the series, but not of the full series because your mean is changing with time, you take a window of data and compute the sample mean and of course, that window length is a user define parameter. So, you take local averages and generate these alpha naught plus alpha 1 k; in this kind of an approach, we do not fit or we do not estimate alpha naught or alpha 1; I just want this component right. So, if you go back to the notation that we have used; I just want this M k; that is all; I have v k being expressed as M k plus wk.

(Refer Slide Time: 16:10)

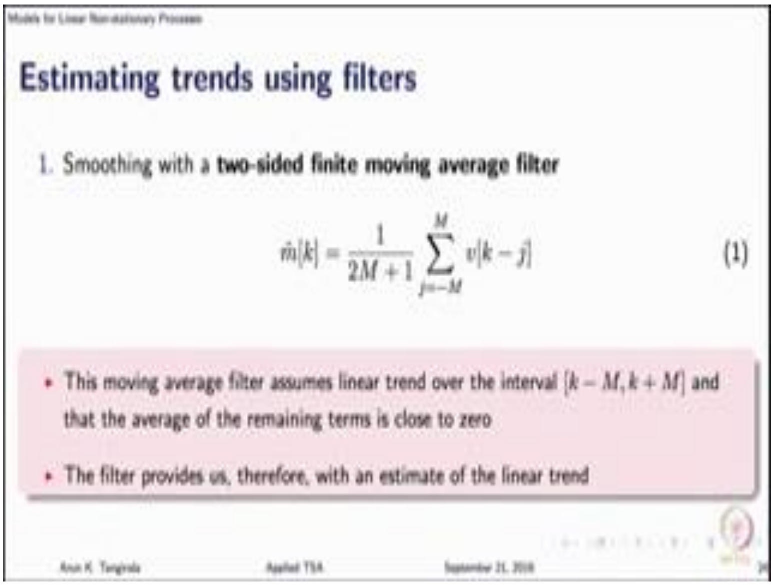
In the first approach, we have assumed the parametric model for M k and then we obtain estimates of alpha naught and alpha 1 or accordingly whatever polynomial assumed. In the second approach, we note that expectation of v k that is the average of v at any instant is M k itself because we have assumed w k to be 0 mean; which means of course, I cannot work with single point therefore, I will work with a small window of the given series and locally estimate M k. So, what you are doing is you are not exactly estimating M at each instant in time.

But you assume that; it is locally invariant, the mean is locally invariant that is an assumption that we are making with; obviously, is not correct, but that is one solution in practice. So, what I do is I construct a weighted average or simple average of the series over a window of data keep moving along and then keep estimating the trend; that is the idea now this average in business amounts to filtering and I do not know how many of your familiar with filtering. Filtering the very name itself suggest that we are eliminating some components of the series; from a frequency domain view point, filtering amounts to only retaining certain frequency components of the series.

When it comes to averaging as you will realize later on; averaging is nothing but a low pass filtering kind of operation that is we allowed the low frequencies to be retained while the high frequency stuff the rapidly change in fluctuation are smoothened out, they are flattened. Naturally when I take a average of any signal that as very you know rapid fluctuation, the net result is the smooth signal which is divide of those fluctuation and those rapid fluctuation belong to high frequency domain regime because there changing very frequently very simple; it just common sense.

So, any averaging operation from a filtering view point is equivalent to let in the signal pass through the low pass filter; low pass filter is one filter which allows only low frequency is go through and that is why this approach that we have where we locally average this series and estimate M k can be thought of as a filtering approach and based on this there are several ideas; this is the one of the ideas here.

(Refer Slide Time: 19:00)



Where you use a two sided moving average filter, so you are standing at; let us say at some point into the series in the initial point; depending on the length of the filter you have to position yourself. You can start from observation one, but then you will not have the left hand side of the data.

So, your series will look like this for example, if there is a linear trend and then maybe there is some oscillation and then on top of it you have these irregularities, as we have discussed the classical approach is to think of the series as being made up of trend plus seasonal plus irregular component and the modern approach is to actually look at it using a multiplicative kind of model; we are looking at the first approach.

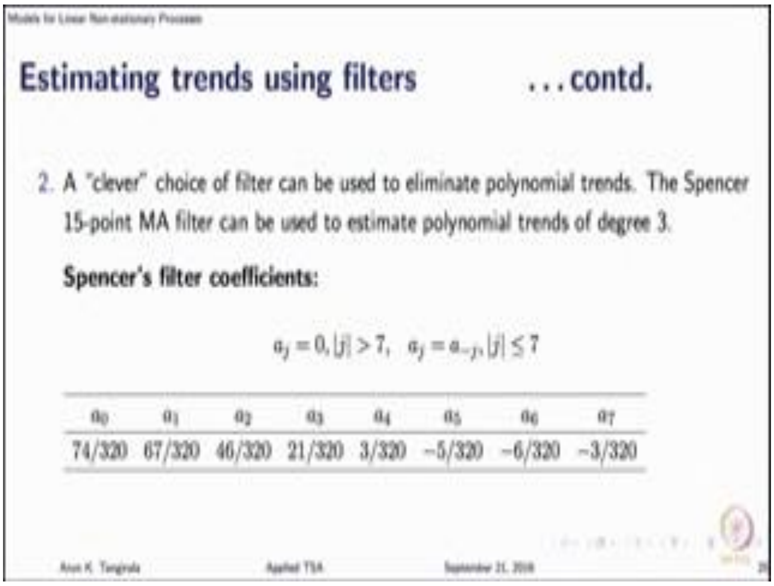
So, if we have linear trend like this what you would do is; using this idea you would construct local averages. So, you would take this window of data and then construct an average and then you move on you; much along in time. When you are standing here, if you are using at two sided filter; that means, that very instant you position yourself and look to the left and right and construct an average.

In the initial points and at the final points, you will have border effects that is standard with any filtering approach you will run into that issue, we also have border issues right even any country will have issues at border also anything will have issue the boarder. When you are coming in the last minute then there is border, when you are submitting assignment at the last minute; you have a border issue.

So, likewise in filtering also at the ends these are called end defects or border effects; invariably you will experience source effects. So, at these points there are some solution, but as you move into the series, the border effects vanish and what is being suggested for example, as I have shown in this screen is to use a simple average but you have equal weights; 1 over 2 M plus 1 where two M plus 1 is the length of the filter, how do I choose M; that you have to play around with; unfortunately this kind of an approach which is a non-linear parametric approach because you are not force fitting any polynomial on to the trend that is why its call the non parametric approach whereas the previous one; force fits a polynomial and that is why its call the parametric approach and you can see there are advantages to both; in the non parametric, you are not pre supposing any polynomial for the trend, but then you have to choose the length of the filter; which is ok.

You do play around with it, there are some recommendation and so on and of course, a simple average may not do a great job because you may want to give more importance to point close to the point that you are located as standing at; suppose you are standing at this point and estimating M, you want to probably give more importance to the points within the vicinity and less importance to the point far off, from where you are position. A weighted average is defiantly better and that is the idea for example, in the spencers method; is not the spencers that we have outside (Refer Time: 22:23) this is a 15 point moving average filter; that means, M is 7 here.

(Refer Slide Time: 22:29)



And the weights are given by Spencers, but by any means you should not thing that this is only filter; there are several other recommendations available for this.

But what you should remember this as a filtering for1, as a filtering approach 2; as a non parametric approach. In contracts to the first approach where we fit a trend; which one should I choose, well it depends sometimes you may have to play around, but sometimes if you see that there is a trend it is better. Parametric approach in general is always good when you have the convection that; yes that model is correct and non parametric approach is good to begin with when you do not have any prior information, you do not know what it is.

For example, you can use this non parametric approach to estimate M and then examine M separately, how M looks like. You may not be able to see that a trend when it is clubbed within the series, with the other component of the series. Once you have estimated M, you can take a relook at M and then hypothesize a certain function of time, then go back and fit a parametric this is standard thing that is followed in any modeling excises. First you being with the non parametric, if you do not have any prior knowledge get some insides and then you use the parametric approach.