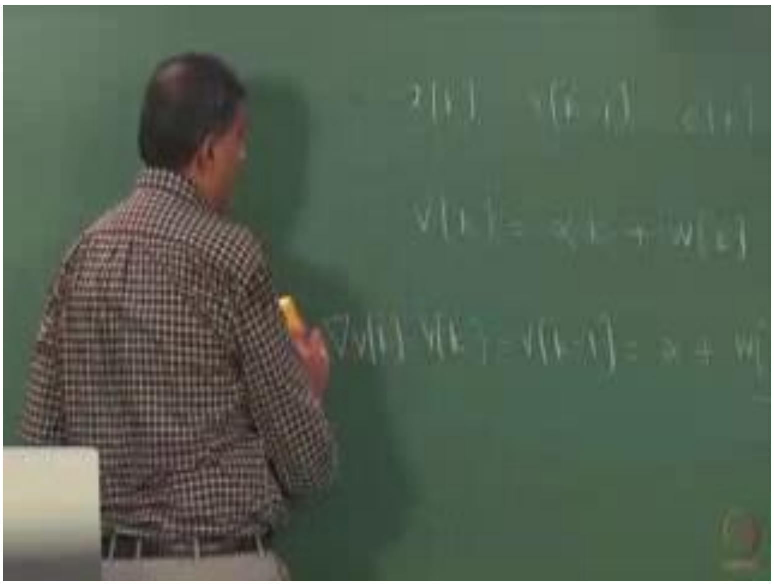
Applied Time-Series Analysis
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture - 27
Lecture 12A - Autocovariance & Autocorrelation Functions-2

Very good morning, before we really get on to today’s class, just a couple of clarifications concerning what we have discussed on stationarity and ergodicity. Yesterday there were a couple of questions after the end of the class. So, for the benefit of everyone I would like to just to discuss those questions may be for a few minutes.

One of the questions was concerning the integrating process and the differencing operation.

(Refer Slide Time: 00:42)



We said that for the integrating process, if x k is an integrating state process, random process then one difference would produce either the white noise or the stationary process depending on whether it is a pure integrating or an ARIMA process.

Now, this differencing operation can not only handle integrating effects what may mean by handle is get rid of the integrating effects, but also can handle trend type non stationarity, remember trend type non stationarities are deterministic type of non stationarities. So, as an example suppose x k had a linear trend and of course, let us say, there is a stationary process riding on this linear trend this is different from on integrating process. So, maybe could distinguish I will have a different notation here if x k was a purely integrating process differencing operations would get you the stationary part which in this case would be white noise.

On the other hand if you have another signal v which is non stationary by virtue of this deterministic trend a differencing operation here as you can see straight away would produce what would you obtain? Alpha plus w k minus w k minus 1, now since w k is stationary, the difference of w k would also be stationary there is no compromise on stationarity there.

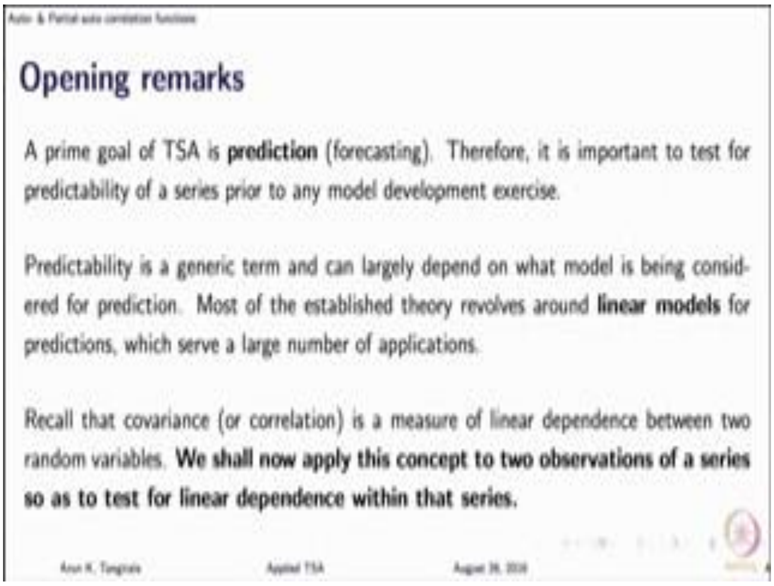
As you can see here, the differenced series which is often denoted by this Nabla here is also now stationary, in other words differencing operations sorry is able to handle or take care of derterminize sorry deterministic trend as well and you can now extend this idea and say double degree of differencing that is a second level differencing can handle quadratic type non stationarities and so on, now that does not mean that you can actually count trend type non stationarities as same as integrating type non stationarities, differencing is just an operation which is able to handle both integrating type non stationarities as well as the deterministic type non stationarities.

In fact, that is one of the reason why the ARIMA models are quite popular because the because one degree or d degree of differencing can candle both integrating effects as well as the deterministic type of trends, but will talk more about this later on I just wanted to give you this clarification that despite the fact that the operation to handle both this trends are identical, does not make the underlined processes identical they have a completely different nature of course, we will also learn different ways of handling this different types of non stationarities differencing is just one of the ways to handle this non stationarities.

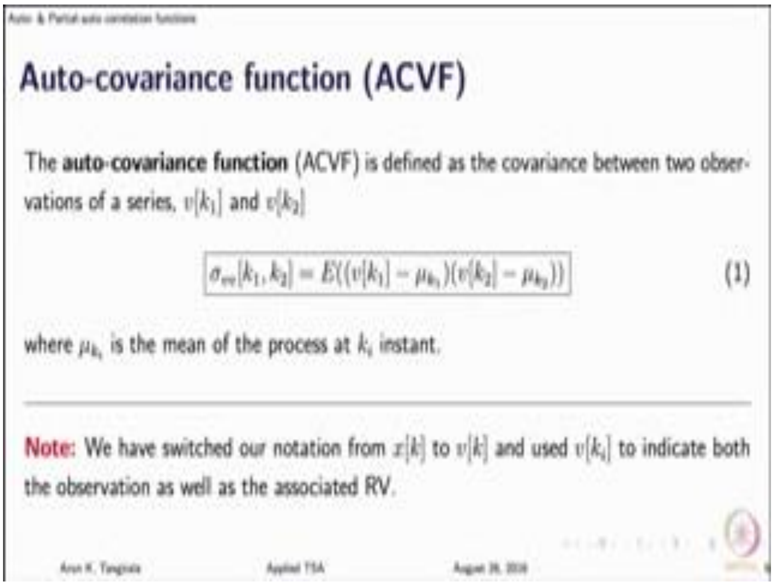
That was a first clarification and the other clarification that I wanted to give you is concerning the second example in ergodicity where we were looking at a constant signal, but that constant is random strictly speaking, it is not a random signal, in the sense of world that is remember we said stochastic process or a random signal one that condition on the past we will not be able to predict accurately, but for that signal alone, if I give the past, you should able to predict the signal perfectly. So, in that sense, it is not a random signal yet, there is some randomness and the randomness is only the initial condition, but that was just an example to highlight the idea of ergodicity, do not take it too seriously, do not take it very close to heart. Now let us move on and get on to the auto correlation and covariance functions and significant part of this course of course, will focus on time series modeling then the other 2 parts we have frequency domain analysis and the estimation theory.

Let us later on to now this main business that we wanted to talk about which is prediction and as I said yesterday the first step in prediction is to test for predictability with a series and we are restricting ourselves to linear models for prediction.

(Refer Slide Time: 05:34)

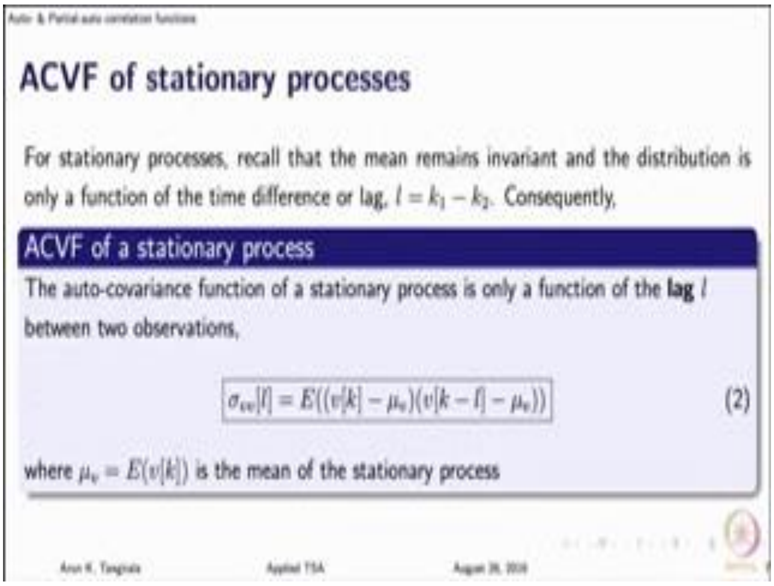


(Refer Slide Time: 05:40)



And for that reason we turn to the covariance function and apply it to any 2 observation of series and then we have the auto covariance function taking bath.

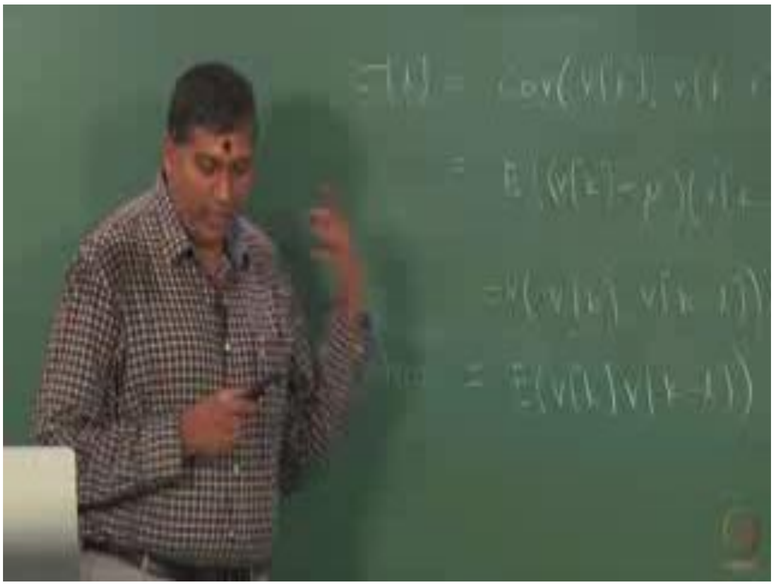
(Refer Slide Time: 05:51)



And for stationary process the auto covariance function is only function of the time difference between the observations not the time stands itself.

Therefore the auto covariance function is the function of the lag the lag is nothing, but the time difference between the observations now there are couple of points that I wanted to make instead mentioned yesterday as well.

(Refer Slide Time: 06:15)



If you look at the definitions of auto covariance, I have not written the subscripts here because in the auto covariance functions, it does not matter, when we talk of cross covariance functions the order of subscript does matter. So, at the moment to save the chalk and my hand ache, I will not write the subscript, but you should understand it all.

The definition of auto covariance is covariance between v k and v k minus l and the covariance itself is defined as you can see on the screen v k minus mu, well expectation of v k minus mu times v k minus l minus, first of all you should straight away recognize that this is an ensemble property it is an average property of the process it is not time average property we are talking of theoretical definitions that is point number one.

Point number 2 is that in many texts you will find the auto covariance mean defined as covariance between v k and v k plus l. So, there is an alternative definition that is used covariance between v k and v k plus l, I am allowed to write at the top, is that ok?

Now, you may ask, is there any difference of course, in the case of auto covariance there is no difference why because it is just a function of the lag its symmetric very very important, not only function of the lag, the fact is that it is symmetric whether you are looking at auto covariance between now and the n l lags on observation l times stands in the past, sorry time difference instance in the past or the influence of the present on an observation that is l times instants into the future, it is same that is what is the nature of the stationarity that we are that we are considering if not strict whites and stationarity.

So, in that sense it does not matter; however, one has to be careful because when you adopt this definition to cross covariance function it can make the difference.

You have to be careful when your reading a text book when you are using any software what definitions has been adopted in the theory or in writing the code because then you can get completely well you know kind of mirror results, but very often one ignores that and gets into trouble.

We will stick to this kind of a definition here even for the cross covariance when we talk about it later on that is point number 2 and the third point which is also equally important is this alternative definition of covariance is used in used in statistical signal processing many statistical signal processing text you will find an alternative definition and a different terminology which is that first you will find the covariance being defined as expectation of v k times v k minus l without any regard for mean centering this is a standard definition that is used in many many statistical signal processing text. So, when you reading a text you should be careful and to confuse you further there is a different term that is used because the purpose of textbook writing is also not to give clarity, but also to give confusion.

Now, this function is called unfortunately the correlation function. In fact, auto correlation function and I know that the statistical signal processing eyes will beat me up for saying that it is unfortunate, but I prefer this definition and this terminology that is followed in pure time series analysis, anyway so each to each his or her own joy, my joy is in using that we will strict to this because you will find this in standard time series text, this definition being used except that sometimes you may find within the time series community v k plus l here. So, do not get confused with this name auto correlation, when in statistical signal processing text, make sure that you go to the definition and see what the author means, so let us move on.

Student: (Refer Time: 11:11).

Sorry.

Student: (Refer Time: 11:18).

What do you think?

Student: (Refer Time: 11:24).

You have the answer with you, you have the question, and you have the answer. So, you can start writing (Refer Time: 11:32) series text books.

Students: (Refer Time: 11:36).

I am sorry.

Student: (Refer Time: 11:40).

It does make a difference in terms of the calculations and for certain signals, it can look different, for example, if you have a constant signal then I mean not constant, but nearly you know very slowly decaying signal it can make a difference and so on. So, in some calculations, it does make a difference let me answer in a more broad sense remember that I showed you an example where there was a where we were establishing the relation between correlation and the coefficients of the linear regression model.

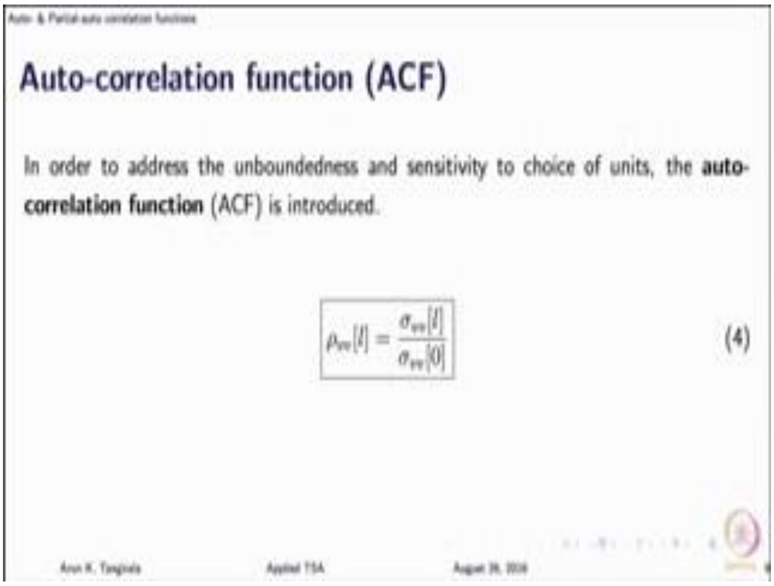
There I showed you that the theoretical relation holds when you consider mean centering. So, in terms of modeling, it is a good question, in terms of modeling, it is almost like I am ignoring the interseptum and that can make a big difference in building linear models for non-linear processes because when you have a non-linear process you never have a linear model, you only have a linearized model and linearized models are models in terms of deviation variables, by not considering the mean you are ignoring the deviations from that. So, in terms of modeling it can make a difference there and so on.

Of course, when the signal is 0 mean, the definition is coincide, but terminology do not coincide, that is a big problem, what we call it has a auto correlation is different from what the others call as auto correlation, but the difference is felt particularly when you are solving the problem and you have to submitted tomorrow. No, seriously it can have you can make an impact for certain class of signals and in developing linear models the you can end up predicting sub optimally; that means, you may miss out the constant term, but then you have to explicitly take care of that in your model, but I will show you at some point in time what difference does it make it is a good question, I will show you by way of simulation, sorry.

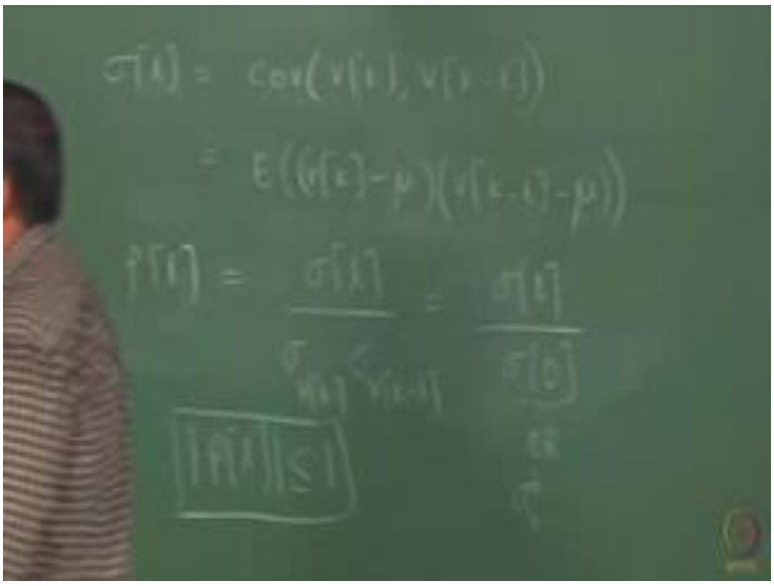
Let us get back to our discussion. So, this is a auto covariance function and as a said yesterday its after all let it heart a covariance function and therefore, you should expect it to have all the DNAs of the covariance correct properties of the covariance, there is no dilution on that it is symmetric with respect to v k and v k minus l, therefore whether you.

It has covariance of v k minus l and v k or what you have written on the board, it is as same and it does not know the physical causation, we cannot tell you which is the cause, which is the effect of course, in a time series we assume that v k minus l is one that influencing v k. So, we do not depend on covariance to tell us which cause the other we are only depending on it to figure out whether its influencing or not that is all and then of course, it is sensitive to the choice of units and so on and therefore, we introduce I am going to skip this, we have already talked about it, will talked about the conforming part later on. And for this the earlier reason that I mentioned which is the sensitivity to the choice of units and the unboundedness, therefore dip to address that issue we work with correlation measures.

(Refer Slide Time: 15:10)



(Refer Slide Time: 15:19)



Now, we have the auto correlation as you see on the screen, this is a row l, sorry sigma l and we have row l which is auto correlation standardized by sigma of v k sigma l divided by product of sigma v k and sigma of v k minus l exactly the same way we worked with correlation again there is no difference here all we are doing is taking the definition of covariance or correlation and applying it to observation of a series.

Now, we know that already the series is stationary therefore, sigma v k should be the same as sigma v k minus l that is the variability that I see a kth instant should be the same as a variability that I see at k minus l therefore, I can write this as sigma l over product of the standard deviations of at any instant in time does not matter that becomes the variance itself and variance is nothing but as we see from the definition, the auto covariance at lag 0, when we write sigma at 0 then it is assumed to be variance. Otherwise you can write sigma square v.

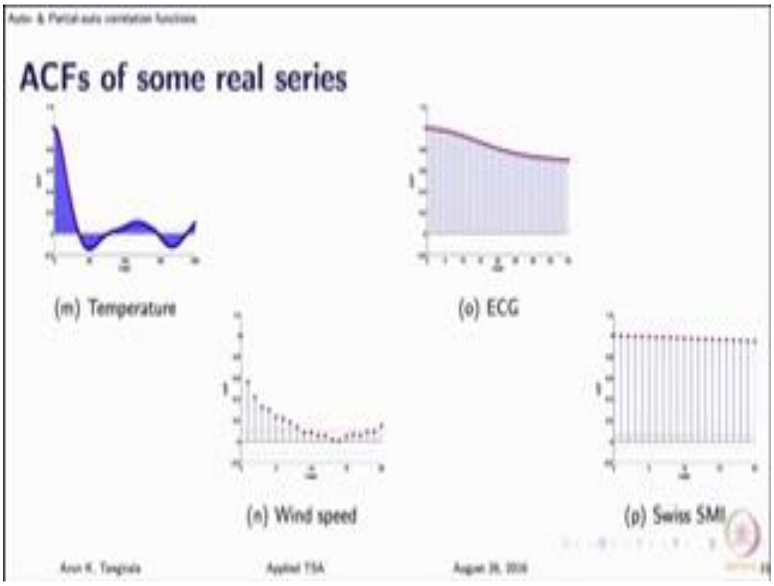
Remember we talked about this notation so you can either write sigma 0 or sigma square v and by definition. Therefore, because it is a correlation measure it is bounded, what do we mean by this auto correlation is always going to be less than or equal to 1 in magnitude and you can see that from the definition as well auto correlation is the auto covariance normalized by the variance at l equal to 0; obviously, it hits a value of one what does it mean the covariance between an observation that is a random variable at kth instant and itself; obviously, should be one because that is the one that can explain the best you if I ask you who looks most like you are the one who looks most like you. So, correlation between you and yourself has to be one better be right.

Now, the correlation between any between 1 observation and any other observation can be at most one that is what this result typically it will be less than one in very peculiar for the peculiar signal maybe it can be one, but otherwise its less than one what this means is; obviously, there is something in one observation that cannot be fully explained by any other observation that is what make a stochastic. So, that is the point to keep in mind of course, as I always say we here right there are seven people who look alike in this world they say I do not know if any of you might have I have not met, but if you do then instantly say that the correlation is1.

But you do not meet any such seven not even one in the stochastic signal each observation looks different from the other at has a different nature to it apart from the stationarity part. So, now, how much an observation is different from one depends on a signal that of course, will study and we are here to exploit such dependence is we hope that the correlation is not 0, if the correlation 0 for between any 2 pair of any pair of observations then we have what is known as a white noise processing and will talk about it shortly. So, I have talk about this now of course, as we have studied in correlation correlation is one, if between any pair of observation if you can express that if that v k minus l is purely a linear function of v k which does not happened for a stochastic signal.

Let us look at the ACFs for some of the series that we have seen in introductory lecture just to get a feel of how ACF looks like, we have not a bottom into the math or the calculations theoretical calculations here, but let us look at some ACF, these are not theoretical ACF, these are estimates and I am not given you the expression for the estimates

(Refer Slide Time: 19:28)



This is the ACF for the temperature series that I showed you, if you recall there are 4 different series that I have showed you remember what the 4 series are?

Temperature and?

Student: (Refer Time: 19:41).

Sorry.

Student: (Refer Time: 19:45).

(Refer Time: 19:47) something that in fact, sorry ECG, very good then one more wind speed; so we had taken 4 differences. In fact, stock market thing reminds me of yesterday when I was getting back home, there is a program that is add on 1 or 2.3 fm on stock market business and so on and callers call in and gets some opinions and there is also discussion on what is happing to India and so on in terms of business. Apparently Goldman’s acts has predicted growth rate of 7.8 percent. So, there was discussion on that as against 7.5 percent big difference now.

There was a discussion and one of the panelists was asked a question, what do you think? Earlier it was 7.5 percent, a Goldman accessed to revised it to 7.8 percent as a forecast, it is a prediction that Goldman acts has made, do you think that makes a big difference while many other families gave some different answers, I like this particular answer because that is practicality this person says look 7.5 or 7.8 ,not a big deal, let us not worry about those numbers because there after all predictions, what matters is a direction and that is exactly what you should keep in mind when you are building time series, do not get so obsessed with numbers, sometimes student gets so obsessed with the numbers that the almost memories, the assignment problem hoping that that is what exactly will turn up in the exam and that is what we call us over fitting in the modeling ARIMA, it is called over fitting.

Over learning; just learn the concept and keep moving, likewise in predictions also what is important is a direction you may be off by 0.3 percent of course, it can make a big difference, but the realities is going to be something different, will wait and see what the growth rate is, I am not into business but I just giving you an example, it just occurred to me and I thought I will share that with you.

Let us get back, here is a ACF of the temperature series and the ACF is always plotted with the ACF on the y axis and the lag on the x axis and because of its symmetricity, we omit the negative lags in the plot as you can see the ACF is taking time to go to 0 and it seems to be not going to 0 at all, since to be oscillator after certain set of lags whereas, the ACF of the wind speed has a different signature, what is the difference that you see? Sorry.

Student: (Refer Time: 22:35).

Always positive, which one is always, is that that is a good difference, I mean it is almost like your readers digest question (Refer Time: 22:45) this 6 differences or any other puzzle type of question, but more than that anything else that you see?

Student: (Refer Time: 22:53).

Ok.

Student: (Refer Time: 22:59).

Is it for the ACF or the series that you are you are giving, you are only expressing ACF functions I was thinking that you probably drawing inferences about the series itself.

Student: (Refer Time: 32:15).

Temperature, periodicity for the series, good what about the wind speed?

Student: (Refer Time: 23:25).

But then what is the purpose of then plotting the ACF, this is the good exercise before going to the theory to like to know how much information you can gather about the series by just looking at the ACFs, remember ACFs is telling us what is the dependence at any lag its telling as what is the influence or dependence whatever way you want to call of one observation over the other right. So, let us think of it as how much the past influences, the present, first of all are these 2 series predictable, everybody agrees it is predictable, why ACF is non zero? At least 1 lag; that means, there is some hope I can use the few observation in the past to make a prediction some prediction right and that is true for both the series.

But as your pointed out the nature of the series is different the nature predictability the underlined function is different even without going through theoretical analysis of the signatures of ACFs, we are able to figure out at least something about the underlined series which we could not have by looking at the series itself right you go back and compare this slide with the slide where you had the series some time you can, but most of the times you cannot drawing inferences by directly looking at the series then you have the ACF estimate for the ECG now this is the completely different one now you kind of understand why you picked this different series because they have completely different characteristics, what can you say about this ECG series?

Student: (Refer Time: 25:05).

Periodic, what you see as for as ACF is concerned? It is almost constant, it is decaying very slowly when you do not know whether the series is linear or non-linear, typically this is a trademark of an integrating process, why because in an integrating process we said s k is s k minus 1 plus some randomness. So, there is a strong similarity between 2 successive observation and the d k is pretty slow that is again feature of an integrating series and then you have the your favorite Swiss stock market index ACF which is even more slowly decaying, it is almost like non decaying, this is again a feature of the integrating process, but one question you want to ask may be some of you have and the back of your mind says is this mapping of signatures of ACFs to the feature of the series unique correct.

I being saying for the ECG and for the SMI that there are going to be there are integrating effects, but did what kind of feature did you see in the SMI, do you recall? How did the stock marketing the next the series look like. So, it was a trend, but its manifesting I am claiming by looking at the ACF alone that it is an integrating effect, do you see the connection between the clarification that I gave in the morning and now that an integrating effect can be actually a trend can manifest has an integrating effect in the ACF domain. So, 2 things we learn one that ACF gives a some idea about the features of the series I mean the underlined kind of functionality when it comes to the predictive relations that are that are present and 2 the mapping may not be unique what you mean by mapping the mapping between the features of the series and the signature of the ACFs.

Now, in all of this, there is one question that hopefully has at least bothered, one of you which is that I have computed this ACF assuming what assumption did I make? Stationary I have use the expression for the estimate of ACF assuming the process to be stationary is the true for the SMI is it is not is the true for the ECG. In fact, it is not we can although you may not see it visibly, but when you go back to the ECG series you can and may be when there are periodicity also its not correct nevertheless I am computing the estimates as for as computing the estimates is concerned I do not need to make any assumptions only when I want to draw inferences about the truth from the estimate then the assumptions taken.

Remember that noting there is noting a legal or you know not something that your violating some law and so on as for as estimating is concerned, estimation is concerned, I can use that estimation for any series whether I can now make valid inference about the original series or not from the estimates depends then only assumptions that I made we will look at all of this, Now one by one first of all will study the theory what is the mapping between a feature of the process or the nature of the process and the ACF signature that is the first thing that will understand then when we talk about the estimates.

We will talk about how estimates themselves behave some lot of times in fact almost all the time the theoretical ACF can look like something and estimates can look like something else. But with the help of estimation theory, we can still make some valid conclusions, so let us move on.