Stochastic Hydrology

Prof. P.P. Mujumdar

Department of Civil Engineering

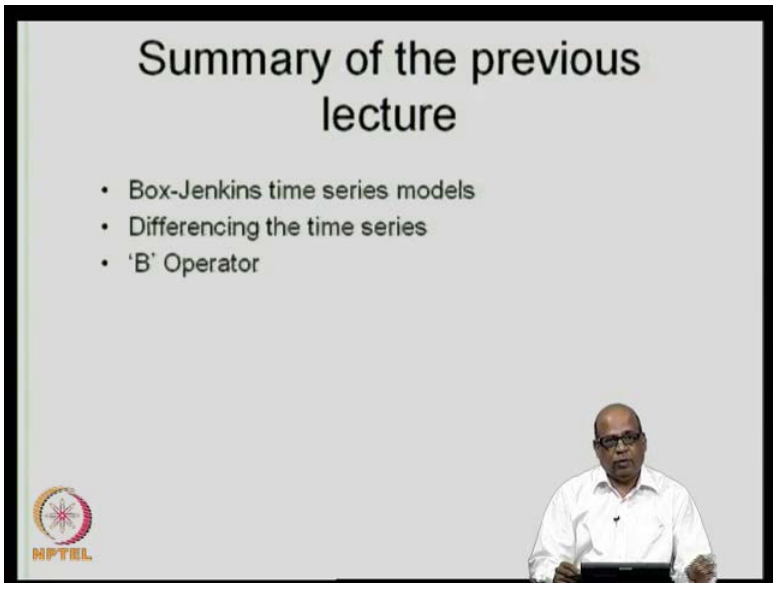
Indian Institute of Science, Bangalore

Lecture No. # 16

ARIMA Models – III

Good morning and welcome to the lecture number 16, of the course stochastic hydrology.

(Refer Slide Time: 00:25)



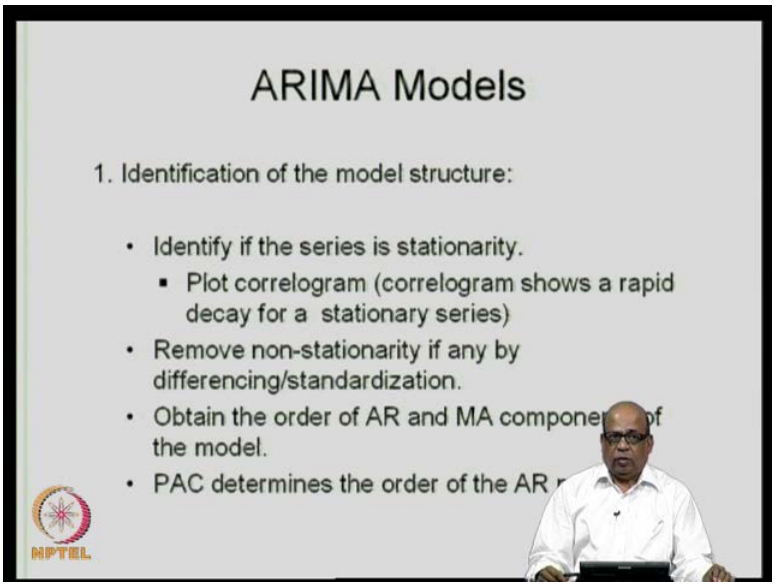
If you recall in the last lecture, we discussed about the Box-Jenkins type of time series models. Specifically, we mentioned about the ARIMA models, the general form of the ARIMA models and then we went on to discuss how to remove the stationarity, specifically the periodicity in the data by differencing the time series. Then we also introduce the B operator by which you can express a time series model in a more compact and elegant form, for example, B of X t, we said is equal to X t minus 1.

Now, in the differencing of the time series, we have discussed the first order differencing, second order differencing and so on, the dth order differencing, in general. So, one minus B to the power d into X t, will give you the dth order differencing as well

as you can recall from the previous lecture. The Box-Jenkins type of models, we mention that there are three major steps that you follow to develop these models. First is identification of the model. You have the time series, major observed data at a particular location, and then you would like to fit a particular ARIMA type of model to this major data.

So, the first step that you do is identify, which among the several models that are possible best to fits the particular data. Now, this we do by examination of the correlogram and the partial or auto correlations, specifically. Then once we identify the model, then we calibrate. In the sense, that we estimate the parameters of the model. Then the third step is the validation of the model itself to examine whether the model that we have so chosen, in fact, satisfies all the assumptions that we have made to begin with, especially, on the residuals or the noise terms as we get in the model. So, in this lecture today, we will go through all these steps of identification of the model, calibration of the parameter estimation of the model and the validation of the model itself.

(Refer Slide Time: 02:51)



So, in the identification, as I mentioned in the previous lecture, first you have to examine whether the series is stationary; time series is stationary, because the model set we are discussing are meant for stationary time series. So, first we check if the time series is stationary and we also said that the correlogram, for a stationary time series, shows a

rapid decay. That is if it is a stationary time series, it shows AR rapid decay. Whereas, for a non-stationary time series, it shows a very slow decay, with respect to the lag, the correlogram drops on very slowly for a non-stationary time series.

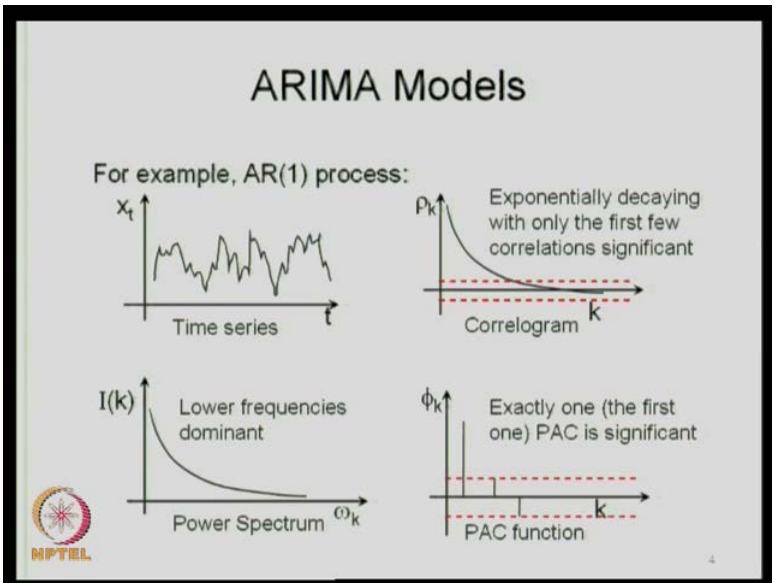
Once we see that the series, the time series is not stationary, then we need to remove any non stationarity that is present. So, that we are dealing with only stationary time series in the ARIMA type of models. Now, how do we remove this non-stationarity? Again, as we discussed in the previous lecture you either difference it, difference a time series X t minus X t minus 1 the first order difference, X t dash minus X t dash minus one for the second order difference and so on.

So, you test the first order difference, second order difference etcetera and then once you do this differencing, again test for the significant of the periodicity or presence of non stationarity in the series so created. Once you are satisfied that the non stationarity has been removed and the series has been transferred into a stationary series, then you start working with the ARIMA models. We then obtain the AR and M A components; that means, how many of AR components and how many of M A components that is necessary for the model. For this we use along with the correlogram, we also use the partial auto correlations.

How to determine the partial auto correlations? I have explained in the earlier classes. Essentially, you use the Yule Walker equation. If you want phi p, let say the partial auto correlation of order p if you want, you formulate the Yule walker equations of order p and determine the AR p. the last phi p that you get in the Yule walker equation in fact, gives you the partial auto correlation for of order p. Now, let us see how specific models look like, in terms of their correlograms and in terms of their partial auto correlations and so on.

What I mean by that is let say you have one theoretical AR 1 process. Let say you write X t is equal to some constant into X t minus one plus some noise term. You generate a series which follows this theoretical model. Plot a correlogram; plot the correlogram of this particular series that you have the generated theoretically from a theoretical model and also the partial auto correlations. Let us see how it looks.

(Refer Slide Time: 06:27)



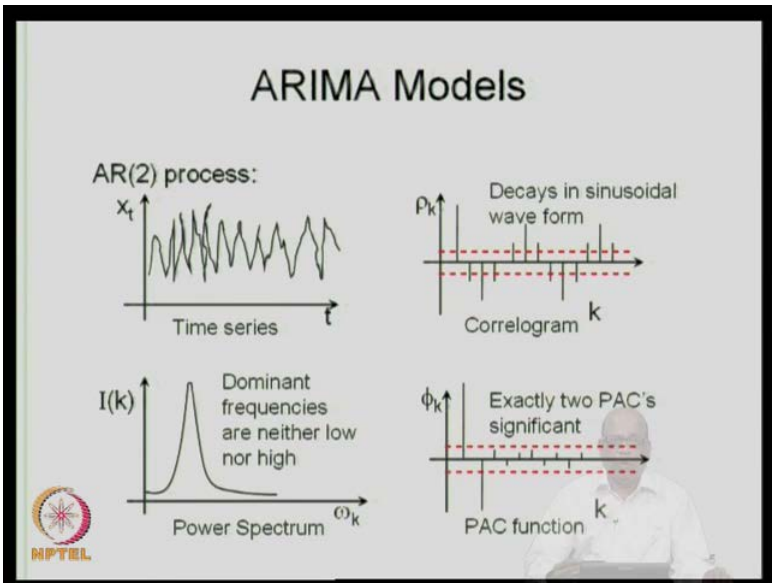
So, if you take an AR 1 model. This is the data that is generated from a theoretical AR 1 model. This is all the data may look like depending on the constant that you use. AR 1, I repeat is written as X t is equal to some constant into X t minus one plus a noise term. So, depending on the constant, you will get some such series. The noise term should have a zero mean and it should be uncorrelated. It should have a The series should be uncorrelated. For this, now, again I repeat that this is data that is generated from a theoretical AR 1 model. For this if you plot the correlogram, the correlogram shows an exponential decay. As you can see here the correlations decay exponentially with respect to the lag K.

So, exponentially decaying with only the first few correlations significant may be about 5 or 6 or only the first few correlations, when you are talking about the large number of lags that you are considering here. The corresponding spectral density looks like this. The lower frequencies are dominant in this particular AR 1 model that we are talking about. The I of K versus omega K that is the power spectrum looks something like this. But, the most important information for AR models comes from the PAC, that is partial auto correlation functions. So, if you plot the partial auto correlation with respect to k, again, recall that the PAC let say phi p. For a particular value of k, let say k is equal to 1, phi 1, you want to determine.

You form the Yule walker equation of order one, you will get phi 1 there. Then by Yule walker equation of order two, you get phi two, order three, you get phi three, and so on. So, you get with respect to k, you get different partial auto correlation. Plot the auto correlation with respect to lag k; you will get PAC function, which is partial auto correlation function. For the AR 1 model, exactly one PAC is significant; in this particular case exactly one PAC is significant.

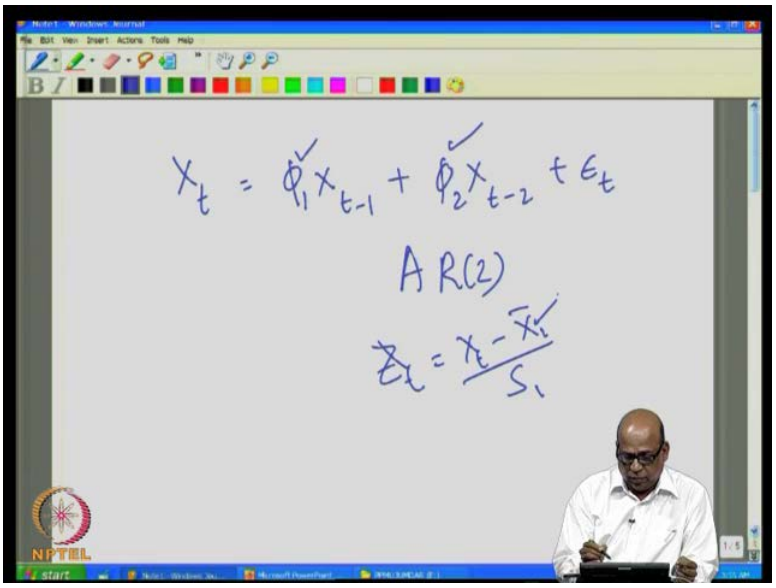
So, there are two important informations, two important features of the AR 1 process that is represented here is that the correlogram shows an exponential decay and exactly one PAC is significant; exactly one partial auto correlation is significant. How do we use this information? Remember, that we have generated this for a theoretical AR 1 process, which means that if the data that you have produces a correlogram, which has an exponential decay which shows an exponential decay and in addition you come up with the PAC function, which has and shows exactly one which is the first one in fact. First PAC is significant.

(Refer Slide Time: 10:10)



Then you can suspect that your time series that you have can be modeled with AR 1process. Let us look at what happens in the case of AR 2 model now. Again, we have taken a theoretical AR 2 process that is we write X t is equal to… how do we write the AR 2 model?

(Refer Slide Time: 10:25)



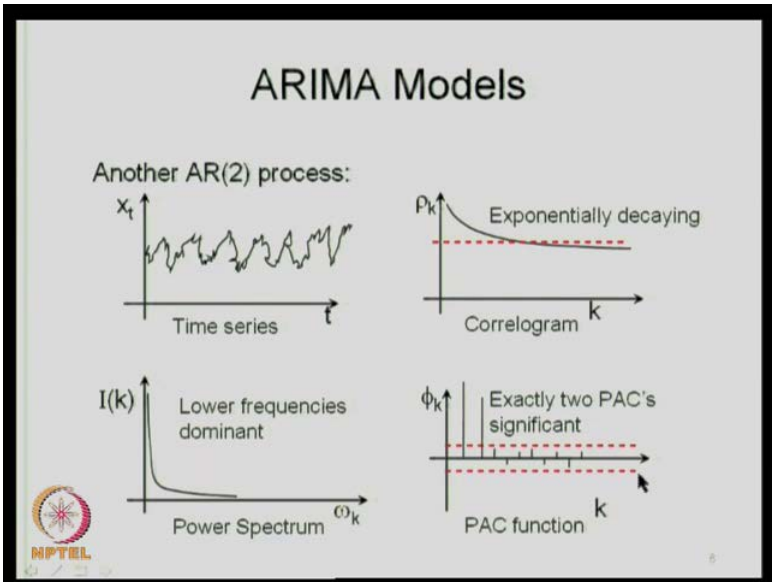
We write X t here for the AR 2 model is equal to phi one X t minus one plus phi two X t minus two plus some error term, noise term epsilon t. So, this is you're AR 2 model. Let say we choose phi 1 and phi 2, if it is a standardize data or some such thing that is between, let say Z t is equal to, we have written Z t is equal to X t minus X bar by S. And if we are talking about the monthly data again, we take those particular months mean and standard deviation to which the time period t belongs.

So, if we write Z t and this AR 2 model in terms of Z t and let us say we choose the constants phi 1 and phi 2 as 0.6 and 0.9 or some such thing. Then plot the time series. The time series looks like this. (Refer Slide Time: 11:38). So, this is the time series that is generated from a theoretical AR 2 model. If we now plot the correlogram, look at what happen in the case of AR 1 process that we used, the correlogram was exponentially decaying. In this case, for AR 2 process as we have generated here, there is decay in the correlogram, all right, but, it is in a sinusoidal manner. So, the correlogram is decaying in a sinusoidal wave form like this. So, slowly that is decay, but it is not an exponential decay, but, it is in a sinusoidal form.

Then we look at the PAC, as I mentioned, the PAC is in fact, the PAC provides the most important information in case of AR processes. The PAC here shows exactly the two PAC's that is the first two PAC's as significant; the PAC corresponding to k is equal to one and the PAC corresponding to two k is equal to two, and there are no other PAC's

which are significant. There are all insignificant. These are the significant bands and I repeat what we mentioned earlier that the same significant band; 95 percent significant bands that we use for correlogram, the same significant band we also use it for the PAC function. What are these? These are plus or minus 2 by root n here, plus 2 by root n here and minus 2 by root n here.

(Refer Slide Time: 13:13)



The power spectrum in this case follow some such shape, where the lower frequencies dominate; however, I should qualify here that the power spectrum corresponding to AR 2 process, different AR 2 processes may behave differently. So, this is not really a feature that we must choose to choose the AR models; however, you are AR model corresponding to,,, I am sorry I will repeat that. The correlogram and the partial auto correlations corresponding to the AR process look like this; for AR 2 process look like this.

We mentioned two AR 2 processes here. The previous one was this. This is a AR 2 process that we generated in which case you have an exponentially decaying and exactly two partial auto correlation significant. This is similar to what we had done for AR 1 process. This is AR 1 process; we have exponentially decaying correlogram and only one PAC significant. For one of the AR 2 models you have the time series like this.

The correlogram decaying, but not in an exponential form, but, in a sinusoidal wave form and exactly two PAC are significant. The dominant frequencies, in this particular case is

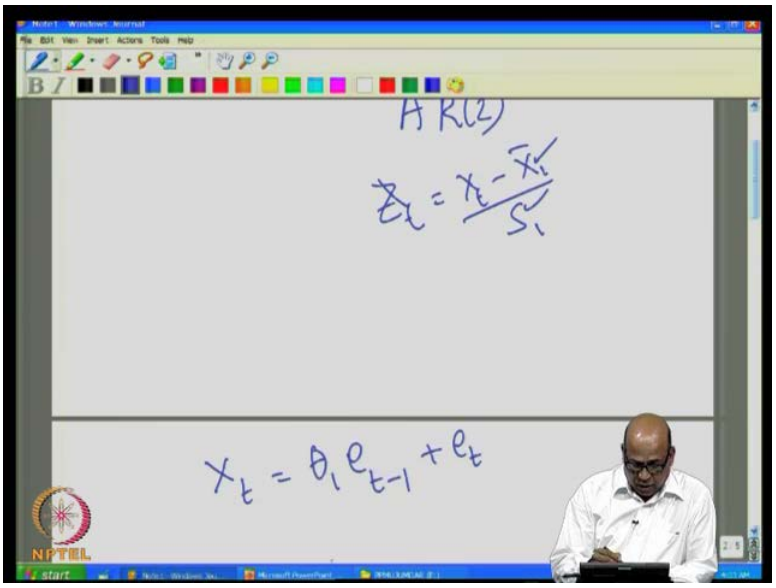
neither high nor low in this power spectrum, but we generate another AR 2 process, by choosing a different set of constant in that phi one X t minus one plus phi two X t minus two plus epsilon t. That is we choose a different set of phi one and phi two and then generate another AR 2 process.

In this case, you get an exponentially decaying correlogram, again exactly two PAC's significant; two partial auto correlations significant. The first two auto correlations are significant; partial auto correlation significant. In this case, the lower frequencies are dominant. So, as I mentioned when choosing AR processes you get the information or indication mostly from the correlogram and the partial auto correlation and not from the power spectrum. The power spectrum can be slightly misleading, but you take the information essentially from the correlogram and the partial auto correlation. So, what did we see? In both the cases of AR 1 as well as AR 2, we see a decay in the correlogram either it is an exponential decay or it is a decay in a sinusoidal wave form.

And exactly one partial auto correlation significant, the first one is significant in the case of AR 1 model. Exactly, two partial auto correlations, which are first two partial auto correlations, are significant in the case of AR 2. So, in general the AR models can be identified by the decay in the auto correlation function. That is the decay may be either an exponential decay, rapid decay or the decay may be in a sinusoidal wave. Slowly, it is decaying, not as slow as I mentioned in the case of non-stationary time series where there is a very slow decay. But, the decay that is seen in a AR 2 model for a stationary time process is it can be in a sinusoidal wave form.

Once you identified that there is decay in the correlogram, then you look at the partial auto correlations. If there is decay in the correlogram, and if there is one or two or three partial auto correlations, which are significant or m number of partial auto correlation significant, then it indicates that the auto regressive parameters are of the order m, either one or two etcetera. So, it is a partial auto correlation that actually gives an idea of the AR terms that you need to include in this. Let us again look at the M A parameters. How does the M A one process behaves?

(Refer Slide Time: 18:18)



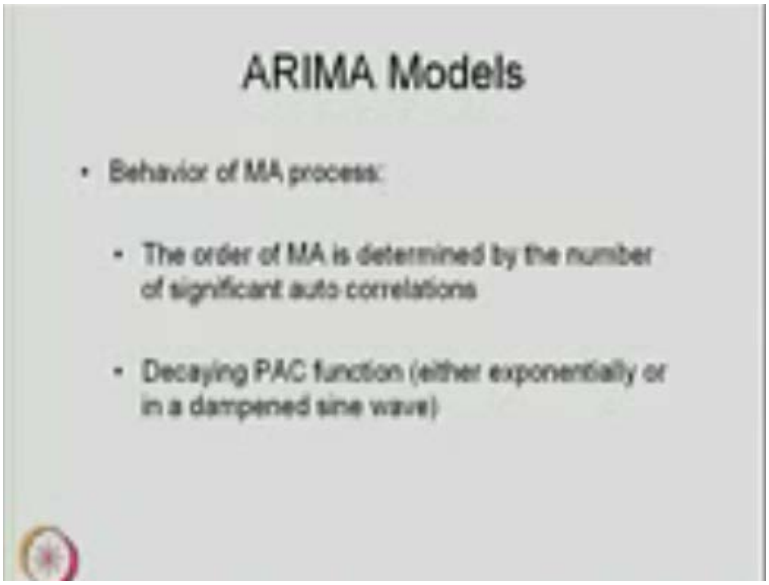
What is a M A 1 process? Again, if we write the M A 1 model here. Let us say what is M A one model? we are writing X t is equal to there is no AR term here. So, phi 1 X t minus one is absent. There is one M A term here, which is I will write theta one and that has to be written with respect to e t theta one e t minus one plus e t. This is our M A 1 model. So, if we write an M A 1 model in a theoretical form like this and then choose a particular value of theta one. So, theoretically we generate the X t value, which follows this particular model and then see how such an M A model performs.

If you look at the M A 1 process, the time series looks like this as we had generated theoretically. X t versus t, the time series looks like this. Look at the correlogram now. The correlogram shows up exactly one auto correlation, which is significant, the first one is significant. The PAC function here it shows a decay in a sinusoidal wave form. So, the PAC function shows decay. The auto correlation shows one significant auto correlation. The first auto correlation significant for a M A one process. Again, the power spectrum is not very helpful in providing us some information. The power spectrum looks something like this. So, for the M A one process for that we have generated with the theoretical M A one model, there is one auto correlation that is significant and the partial auto correlation decays slowly.

Let us look at the M A two process now. Again, what we are doing here? X t is equal to theta two theta one e t minus one, plus theta two e t minus two plus e t, that is the model

we are taking here. The correlogram shows exactly two auto correlations; the first auto two auto correlations significant and the PAC function again shows a decay in a sinusoidal wave form. We can generate another M A two process similar to what we did with the AR 2 process by changing the theta 1 and theta 2 and again check that there are exactly two auto correlation functions and there is a decay in the PAC function. Now, the decay may be in different forms. It may be either in an exponential form; exponential decay form or in the terms in a sinusoidal wave form in the pre PAC.

(Refer Slide Time: 21:13)

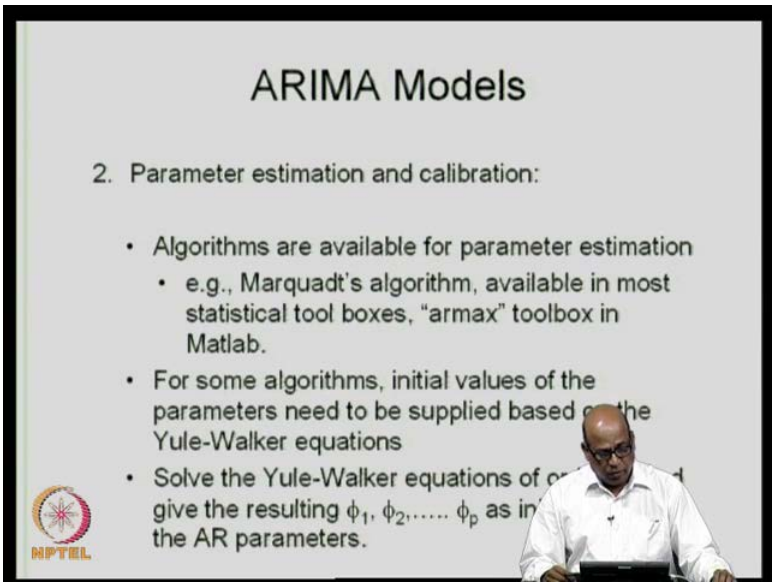


So, the M A models behave as follows. If there is decay in the PAC functions, either exponentially or in a dampened sine wave form, then you suspect that there is an M A process that is in play there, in that particular process, then you look at the auto correlogram. So, if your PAC function shows the decay, then you look at the auto correlation function. The order of M A is determined by the number of significant auto correlations.

Now, this is how you fix the AR parameters and the M A parameters. As you can readily see, if it is a purely AR model then we know how to do it. If there is a if it is a purely M A model then we know how to go about this; that means, if it is a purely M A model you know that there is a decaying PAC function and then you look at the auto correlations of first or second auto correlations which are significant and so on. But, when we have both AR as well as M A process is together that we are looking at ARIMA model, then this

identification, this procedure of identification becomes slightly complicated and it does not always lead to unique identification or unique model identification.

(Refer Slide Time: 22:44)



In fact, with the advances in the computations and so on, the identification of the models has become slightly, I would say not necessary, in the sense that you can choose several candidate models and for each of this candidate models, you examine how, whether the data that you have fits any of this candidate models well or not. In fact, in the case studies that I will be discussing we will discuss that particular methodology also.

So, assuming now that we have identified, which particular model fits the data that you have reasonably well. Then the next procedure is parameter estimation. Let us say that we have identified that you have one parameter for AR and one parameter for M A theta one and that is phi one and theta one. So, we need to know the estimate of the parameters phi one and theta one. There are large number algorithms we do not have to go into the developments of the algorithms itself, but there are large number of algorithms available for estimation of parameters of ARIMA models.

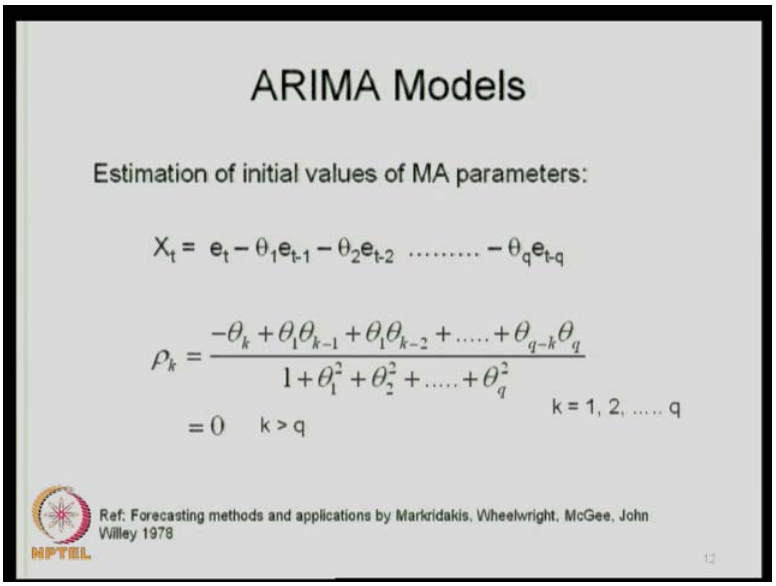
The most commonly used algorithm is the Marquardt algorithm, which is available in most of the statistical tool boxes, but we can also use tool boxes in the matlab, where the "armax" toolbox gives for ARIMA models. It gives the parameters directly. If you are using algorithm such as Marquadt's algorithm and you want to write your own codes, for that it typically requires some initial values of the parameters. That is let say that you

have a ARIMA two one model, where you have two of AR parameters and one of M A parameter to be estimated. Then you have to give the initial the algorithm is such that it starts with an initial assumed values and then it slowly convert this to the correct values, the correct estimated values.

In such cases, the AR parameters, if you want to give initial values initial guesses for the AR parameters, you can simply solve the Yule-Walker equations. Let say you have p AR parameters, simply, solve the Yule-Walker equations of order p and give all phi one phi two etcetera, phi p. Remember, that phi one, phi two etcetera are not partial auto correlation; only the phi p is a partial auto correlation. But, you give phi 1, phi 2, phi p etcetera as initial guesses to the Marquand's algorithm. Then the algorithm proceeds starting with these initial guesses and then gives you the final converts values of parameters that are for AR parameters.

Let see you want to give a guesses for M A parameters. There are certain algorithms, I would repeat that if you want to use toolboxes in mat lab, for example, armax toolbox in matlab, and then there are no initial guesses necessary. But, there are some algorithms, which will require that you provide the initial guesses, in which case you can use the Yule-Walker equations for providing initial guesses of the AR parameters.

(Refer Slide Time: 26:20)

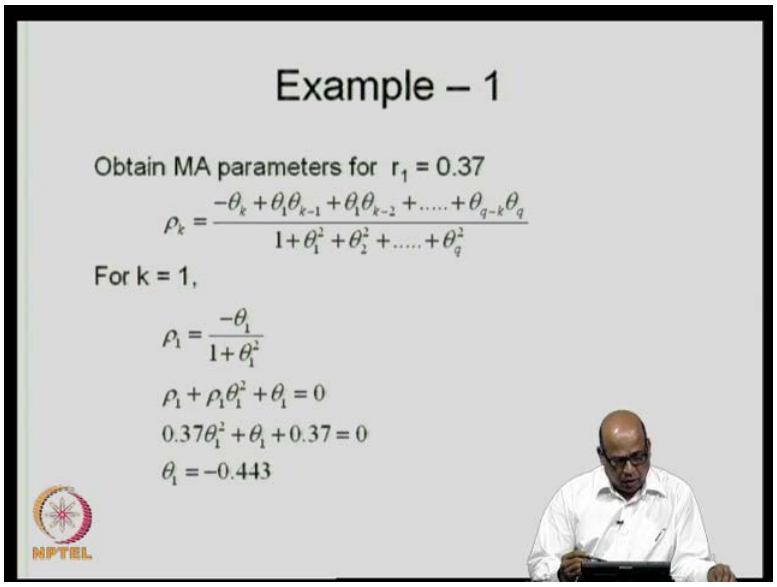


And for the M A parameters, if you want to give an initial guess, you use this equation this is from the textbook forecasting methods and applications of John Willey, 1978. I

have taken this from this particular book in which you write the M A model like this, X t is equal to e t minus theta one e t minus one etcetera. So, all the thetas are negative here. In our convention we have put all the thetas as positive.

So, if you use this particular form then you can estimate the theta's that is a initial values of theta's to be provided to the algorithms to get the correct estimated values of theta's. you can use this rho k is equal to minus theta k plus theta one theta k minus one plus theta one theta k minus two plus etcetera plus theta q minus k theta q; k is equal to one to q. What is q? The q is the order of M A terms. That is you have q as M A terms. In this model M A q model is what we are talking about. So, for k greater than q you have rho k is equal to zero.

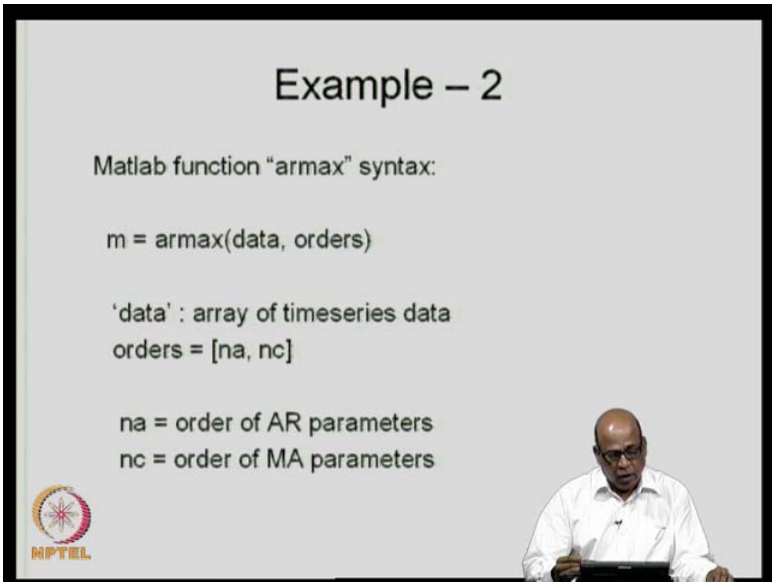
(Refer Slide Time: 27:36)



Now, let say that rho k is auto correlation at lag k. So, we write this expression and let us say we want to obtain theta one corresponding to r 1 is equal 0.37; r 1 is the sample estimate of rho 1, which is auto correlation at lag one. So, we want to use this expression to provide an initial guess for the M A parameter, to be used in an algorithm set as Marquand's algorithm, for getting the M A parameters. So, we write this k is equal to one. So, rho 1 is equal to minus theta one. There are no other terms here, divided by 1plus theta 1 square that is all. So, rho one is equal to minus theta 1 by 1 plus theta 1 square from this simplification you get theta one is equal to minus 0.443.

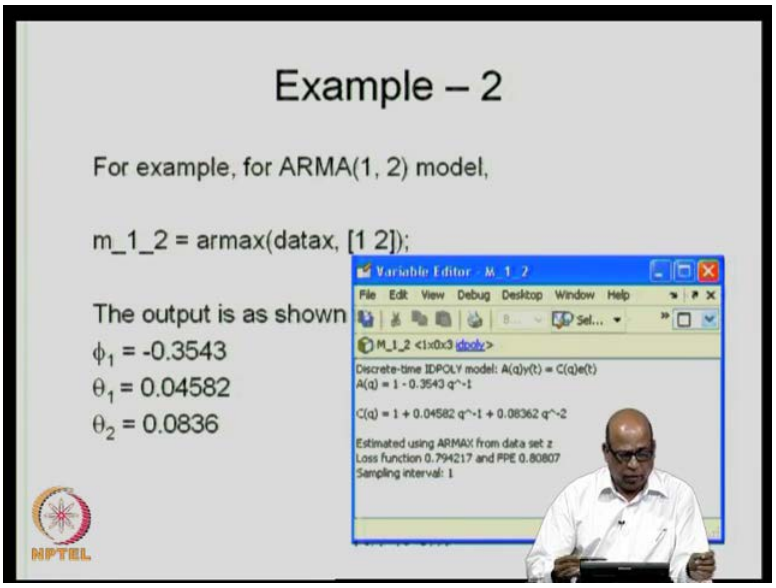
In fact, if you do some numerical experiment like this; that means, that you provide the initial guesses into the Marquand's algorithm and then see what is the final value of theta one that you get. If you provide a completely different initial guess, let say you provide instead of minus 0.443, you provide plus 4.5. Even then you will get through the algorithm thus more or less the same value as you would have got with this initial guess. Except that the time that it takes to converge to that value will be higher. So, you need not unduly worry about providing the right initial guesses. The right initial guesses will only save you the computational time, but the accuracy of the final result will not be so much affected by your initial guesses.

(Refer Slide Time: 29:36)



So, if you want to use armax for example and this is what many of our students use when we give assignments, they use the matlab functions. So, if you have access to mat lab, you can use the armax syntax, armax directly from the matlab. This is simply you say any variable that you want to set, is equal to armax data. This is a array of the time series data and orders. An order is you will provide the number of AR parameters and number of M A parameters. By this you are specifying the time series the complete time series and here you are providing the order of your AR and M A terms.

(Refer Slide Time: 30:29)



Say for example, you want to use or want to estimate parameters for ARMA (1, 2) model. So, I define this as my result m 1 2. I did define this as result is equal to armax. Those who use to mat lab can readily see this. Amax, this is a function call datax, I have defined the series of data in this array datax and then one coma two, it says that AR parameter is one M A parameter is two. Then we get the output like this, which indicates phi one, remember what did we write for AR term using the B operator, it is one minus B into X t. So, that is the form in which it is giving that is one minus phi 1 B into X t. So, phi 1 is minus 0.3543.

Then similarly, for M A 2 parameter we have written you can just verify from the previous lecture. How do we write? One plus B into theta one, into X t plus B into that is B square into X t. So, this is you get theta two as 0.0836 here and theta one as 0.04582 and phi one is this. So, this is how you can use the built in tool boxes in matlab and several other statistical programs that you may have access to. Most of the statistical programs will have some functional calls, which will return the parameters of a general ARMA model. So, we will use those. I do not want to go into the details how the algorithm themselves work in this course. We will just see how to use this different tool boxes get the parameters. Our focus here is once we get the parameters what do we do once we get the model structure what do we do.

So, we have completed two steps now, one is identification of the model structure. How many AR terms, how many M A terms and what is the level of differencing that we will like to do? The level of differencing, the order of differencing, that we want to do will depend on whether you are able to convert the time series into stationary time series or not. Then you will fix your AR components and M A components. As I mentioned for higher order terms like AR five and then M A three, M A four, etcetera, if you want to have a combine models, let say you are talking about ARMA 6, 5, then the identification procedure becomes quite cumbersome, quite difficult and perhaps you may not get a correct identification at all.

Typically, what we do is especially in the most hydrologic applications, rather than identifying specific AR and M A terms, we formulate several candidate models. Let us say that I will go from AR component one, two three, four etcetera up to six. Similarly, M A, I go from 1 2 3 etcetera. So, we will formulate ARMA 1 1, ARMA 1 2, ARMA 2 1, ARMA 2 2 etcetera, like this. We formulate several candidate models. Then on these candidate models, we do the parameter estimation. For each of this candidate models, we send it into let say armax and then corresponding to each of this model, we calculate the parameters, which means that we have now formulated all the candidate models complete with the parameters.

Then among these candidate models, which do we choose, which among this candidate models are so considered, will fit our data the best. That question, we answer subsequently by either maximum likelihood or minimum mean square error criteria and so on. So, essentially while there are procedures available to identify the AR terms and the M A terms separately, these procedures are not really essential when we are doing applications, especially in hydrology, because the AR terms may go from one to four or one to five and so on. We hardly get AR terms of 20 and 30 and so on, especially when we are looking at monthly stream flow series, seasonal rain flow series and so on, whatever we generally get in the hydrologic applications.

Similarly, the M A terms, we hardly go beyond two or three. it is either one two or maximum three we may go. Also there is a concept of contiguous and non-contiguous models. By contiguous, I mean let say you are talking about AR 2 model. AR 2 model will necessarily have phi one X t minus one plus phi two X t minus two, plus error term,

noise term. This is a contiguous model in the sense that when you are talking about AR 2 both second that is t minus 2 and t minus one, both are involved.

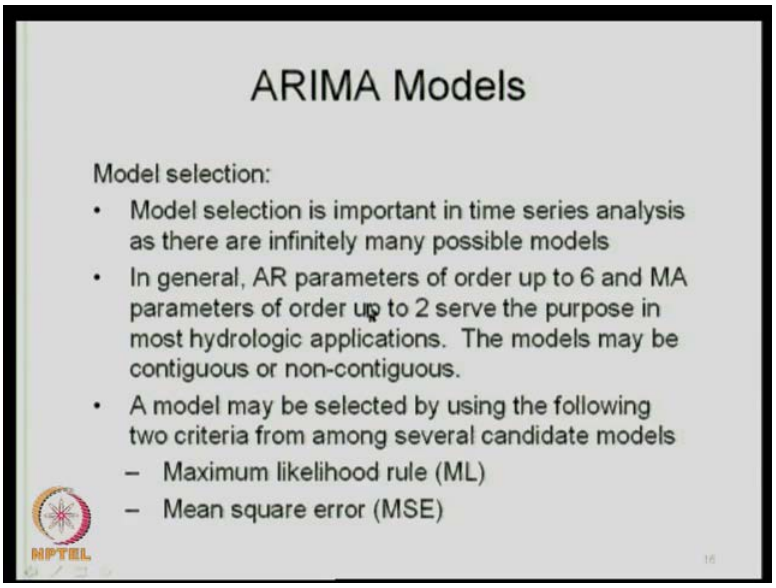
Whereas a non-contiguous AR 2 model, I can write it as it will have two parameters, only two AR parameters only, but it need not necessarily have t minus 1 and t minus 2. It may have t minus 1 and t minus 12. So, I may write AR 2 as X t is equal to phi 1 X t minus 1 plus phi 2 X t minus 12 plus error term of course. So, like this you may construct a contiguous model and or a non-contiguous model, depending on significance of the correlations that you come across the data and so on.

When we explain or when we deal with the applications this point would be clear. The point that I would like to drive home is that the estimation procedure, while there is an elegant way of doing is rather subjective. But in hydrologic application, we do not have to be unduly concerned about it; we can choose candidate models appropriately with proper judgment. For all this candidate models, you apply the validation test, do the parameter estimation and then apply the validation test, and then you will be able to fit the model correctly.

All right now, as I mentioned, you know virtually there are infinite number of model that are possible. You have a time series, you have a observed time series, you can formulate theoretically virtually infinite number of models, because by varying p d q, theoretically there are large number of models that you can formulate, but it is important for us to choose among these several models that are possible, choose the best model that is that well represent the data.

Remember here, all this exercise we are doing so that the data that we have observed can be represent can be replaced by a time series model of ARMA type, so that this model can be use to generate data for actual applications. So, you want to build a reservoir and you want to generate data using this kind of models and therefore, the decisions that you are making on the capacity of the reservoir or the spillway capacity and so on, all of these will depend on what kind of sequence that you have generated and therefore, a wrong model will always be not only expensive, it can be quite risky also, if you generate data with the wrong model. Therefore, it is important that you select the best model the most appropriate model for the data.

(Refer Slide Time: 39:05)



As I mentioned, in general, in hydrology we do not go beyond about six. So, AR parameters of order up to about six and M A parameter of order up to about two, serve the purpose in most hydrologic applications. And again the models can be either contiguous or non-contiguous. So, you may still stick to, let say AR 4 or AR 5 model, but not all the AR terms be contiguous. X t minus five, X t minus four, X t minus three etcetera, you need not go in that order. You may choose based on your initial analysis on the correlations and a significant period citizen and so on.

For what I meant by that is let say you are talking about monthly stream flows; obviously, monthly stream flows, the immediate previous stream flow, has a strong influence on this particular month stream flow. So, X t minus one has to be there, but whether the stream flow, two months behind, is influencing this as much as the stream flow one year behind influences this. For example, the June months flow of this particular year may have a strong correlation, stronger correlation with the June month flow of the previous year, compare to its correlation with the May, with the April month flow of this year.

So, we may write an AR 2 model, as X t is equal to phi 1 X t minus 1, which takes care of the may month plus phi 2 into X t minus 12, of one year behind. So, you are looking at June of the previous year and therefore, this becomes a non-contiguous model. By doing this, what we are doing is we are avoiding large number of terms. Suppose you wanted to

put X t minus 12 also in the model and you wanted to build a contiguous model. What would it look like? It will look like X t is equal to phi 1 X t minus 1, phi 2 X t minus 2 etcetera plus phi twelve X t minus twelve.

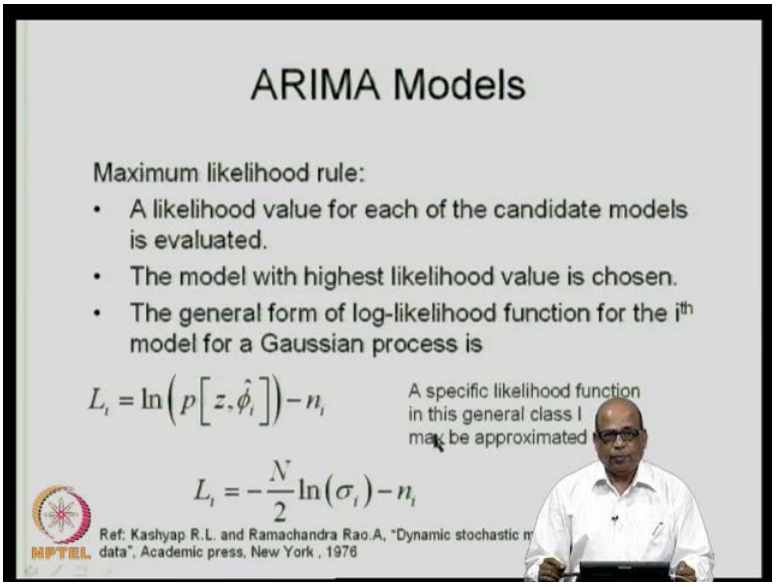
All the terms have been included and therefore, the number of parameters that you need to estimate will be quite large, phi 1 phi 2 phi three etcetera phi twelve you have to estimate. Whereas a non-contiguous model, you would have put X t is equal to phi 1 X t minus 1 plus phi 2 X t minus 12 directly plus the error term. So, that is an advantage of using the non-contiguous models. Now, once you have such collection of models, these are the candidate models, now with these candidate models now, we need to choose which among these models best fit our requirement.

We use two rules here one is the maximum likelihood rule the other one is called as a prediction method or the mean square error method. So, we choose the models based on either maximum likelihood or the minimum mean square error. The model that we are so building can be used for two primary purposes. One is generation of long term data; that means, from the historical data, you have built these particular models, use this particular model to generate the large sequence of data. Let say from the fifty years data that you had, you will generate another fifty years data, another hundred years of data, several such sequences and so on.

That is for generation of the data whenever we are talking about generation of long data series, and then we use the maximum likelihood rule to choose the models. The other purpose that we will be using these models is for short term forecasting itself. Let us say we are standing in the month of June, you would like to use this particular model to forecast what is likely going to happen in the month of July. This is called as one time step ahead forecasting. In the two time step ahead forecasting, standing in the month of June, you want to forecast what is likely to happen in July, as well as in August two time step ahead forecasting. So, you are essentially looking at shorter time duration of for use of the models. So, one is long time generation or synthesis of data. Another is short term forecasting of the data. For long term synthesis of the data, the criteria that we used to select the model are typically the maximum likelihood criteria, and for the short term forecasting problem, we use the minimum mean square error data criteria. It is also called prediction approach.

Classically, there is another criterion, which is called as Akanke information criteria, AIC criteria, which also can be use to select models. But, Kashyap and Rao, the textbook that I keep referring in the due course, they have recommended these procedures; because they have found the AIC that is Akanke information criteria not exactly suitable in many situations.

(Refer Slide Time: 44:50)



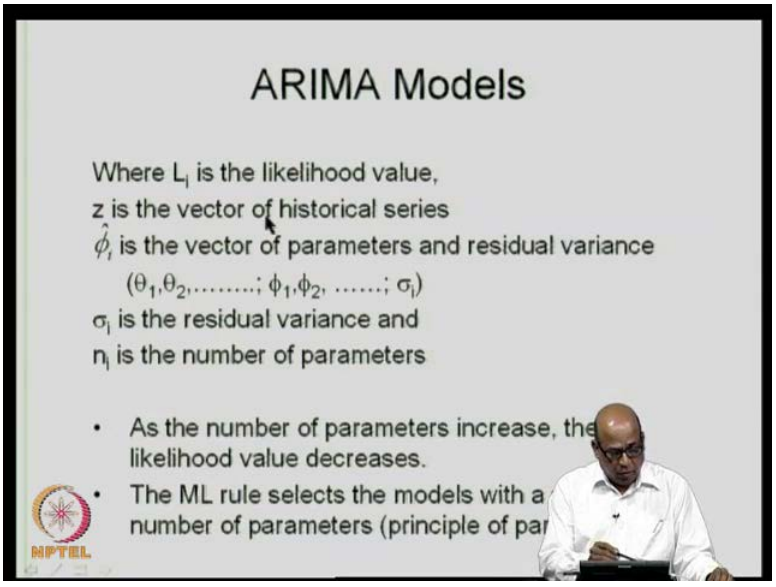
What does the maximum likelihood rule look like now? You remember what we did in our parameter estimation using maximum likelihood method. we formulate a likelihood function and then look at the maximum value of that. Essentially, the principal is same here, but we calculate the likelihood values for individual model. Remember, that we have formulated candidate models. Let say there are ten candidate models. For each of the candidate model, we formulate we determine the likelihood value of the candidate model and then from among this candidate models, ten candidate models, we pickup that particular model which gives us the maximum of such likelihood values.

So, we formulate a likelihood function for each of the candidate models. Then we choose the models with the highest likelihood value. This is the book I am referring to Kashyap and Rao, 1976, Dynamic Stochastic model from empirical data. It a classic book in time series analysis. They provide a general form of the likelihood function. This is a log likelihood function, L i is equal to log p z phi i. This is a probability distribution minus n

i, the n i here is the number of parameters. Let us say you have ARMA 2 1, the total number of parameter is three; n i is the total number of parameter.

Now, this is a general class of the log likelihood functions. A specific likelihood function in this general class may be written after some approximation as L i is equal to minus N by 2 log sigma i minus n i. This capital N here is the number of data values and this is the standard deviation of the residuals and this is a number of parameters.

(Refer Slide Time: 47:04)

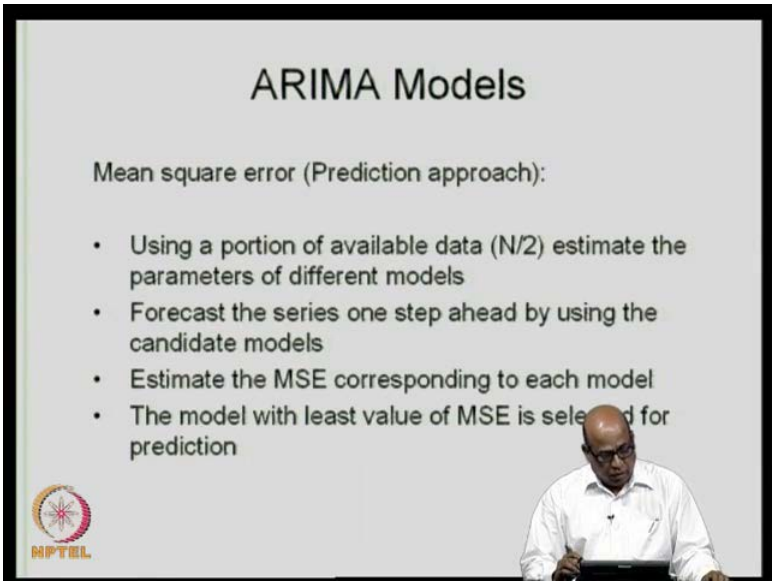


So, L i is the likelihood value and if you want to understand this general form, this is your vector of historical series, Z is historical series. The phi i cap is the vector of parameters and residual variance. That is sigma i is the residential variance. So, this is in this particular form where is in sigma i here. And n i is a number of parameters. Look at this, which will be using to select the models. We are choosing the model those results in the maximum likelihood. For each of the model, i here refers to it model.

So, for each of the model, we are calculating the likelihood function and n is the data, sigma i your standard deviation of the residuals, and n i is the number of parameters as your number of parameter increases the likelihood value decreases let us say from ARMA 2 1 ARMA 3 1 in general, in general it also depend on your sigma n, for the same n it also depends on sigma i. But, for the same sigma i, as yours number of parameter increases then the likelihood function decreases.

So, the likely maximum likelihood rule selects the models with a small number of parameters, because the number of parameters as it increases the likelihood value decreases and therefore, it is penalized in a way. Now, this is called also as the Principle of Parsimony that is generally valid in Box-Jenkins models; that means, from among those many models that all may represent your data reasonably well to choose that particular model, which has the least number of parameters, because for a given data length estimation of parameter, becomes an issue estimation of parameter, becomes an issue and therefore, you should have a model with as few number of parameters as possible. Now, this may be counter intuitive slightly, because we may think that the more complex models with more number of parameters may perform better compare to simpler models. But the Principle of Parsimony states that you must go with models, which have lower number of parameter, less number of parameters.

(Refer Slide Time: 49:50)



Now, this is for the maximum likelihood approach. So, what we do in maximum likelihood approach I will summarize here is that we choose a number of candidate models let say ARIMA 1 1, ARIMA 1 2, ARIMA 2 1, ARIMA 2 2 etcetera. Like this we choose a number of candidate models. For each of the models, we estimates the parameters phi 1 theta 1 phi 1, phi 2 theta 1 theta 2, etcetera like this for each of. So, candidate models you estimate the parameters.

So, all the models are now ready. Once these models are ready now, you estimate the likelihood values for each of the models. This L i is for a particular model. So, for each of the model you know the number of parameters, you know the residual error that you are getting out of that parameter, I am sorry, the residual standard deviation associated with the model i.

Now, how do we get this? We get this by applying the model on part of the data and you get the residuals out of that and then get the standard division of that. Through numerical example, it will be clearer and you know the n therefore, corresponding to each of the models, you get the likelihood value. So, for all the ten models now, you have the parameters, you have the model structure, you have the residual series and you have the likelihood values. Then you choose that particular model with results in the maximum likelihood. So, that is the maximum likelihood criteria, which is generally used for long term simulation of data. If you want to use the long term simulation of data and that is a procedure that you generally use.

Now, for the short term forecasting problem, one time step ahead or two time step ahead forecasting etcetera, which are extremely important in hydrologic applications, where you, let us say you are looking at flows into the reservoir over the next month or the over the ten days, so that you can operate the reservoir in real time. So, such short term forecasting problems are extremely important in hydrologic applications, both for normal operation requirements, also for a extreme event forecasting requirements. For example, you may want to forecast the flood during the extra time period and so on.

So, if you want to do such forecasting then the type of model selection procedure that we adopt will be different from what we did for long term simulation of the data. So, that is what we do in the minimum mean square error approach or the prediction approach, where you start calculating using the model, what is an error that you get and then sum the errors and then get the mean square error, choose that particular model which results in the least and lowest error.

Now, this procedure we will discuss in the next class. So, essentially what we have covered in this lecture is starting with the expression of ARIMA models using the B operator. We saw the process of identification of AR and M A terms. Recall that if your auto correlation function that decays either in a exponential form or in a sinusoidal form

and you have one or two or three partial auto correlation significant the first, second, third, partial auto correlation significant. the number of partial auto correlations that are significant will indicate the order of AR. Now, in the M A it is slight, it is exactly reverse that is in the M A identification that is a moving average identification, the partial auto correlations decay either in a sinusoidal way or in an exponential way. You have one or two or three significant auto correlations corresponding to M A 1, M A 2 or M A 3. This is how we identify AR and M A term.

However, as I mentioned, we generally adopt, we formulate candidate models, large number of candidate models and from among those candidate models we choose which is the best one for the data that we have and for the purpose for which we want to use. We introduce the maximum likelihood criteria for selection of the models. We have also discussed how the parameters are estimated for each of these models ARMA models, using algorithms and typically the armax function that is available in mat lab can be used straight forward. All we need is your data series and the models structure how many AR parameters and how many M A parameters.

So, you formulate large number of candidate models, estimate the parameters, and then we calculate the likelihood function for each of, sorry, the likelihood value, I am sorry likelihood value for each of the models and pickup that particular model, which results in the maximum likelihood value and for the short term forecasting. We use the method of minimum mean square error.

Now, this method we will discuss in the next class. Essentially, what we do there is that let say that you are talking about AR 1 model for short term forecasting, you have estimated the parameter and then you apply that particular model for half the length of data n by two, in estimating the parameters, and the remaining n by 2 values, you choose to get the errors in the forecast; that means, you forecast one time step ahead, get the error and accumulate all this errors, squares of the error and get the minimum mean square. So, like this you do this exercise for all the models, you will get the mean square error corresponding to all the values, all the models. Choose that particular model, which results in the minimum mean square error that is what we do for forecasting purposes. So, we will continue this discussion in the next lecture. Thank you for your attention.