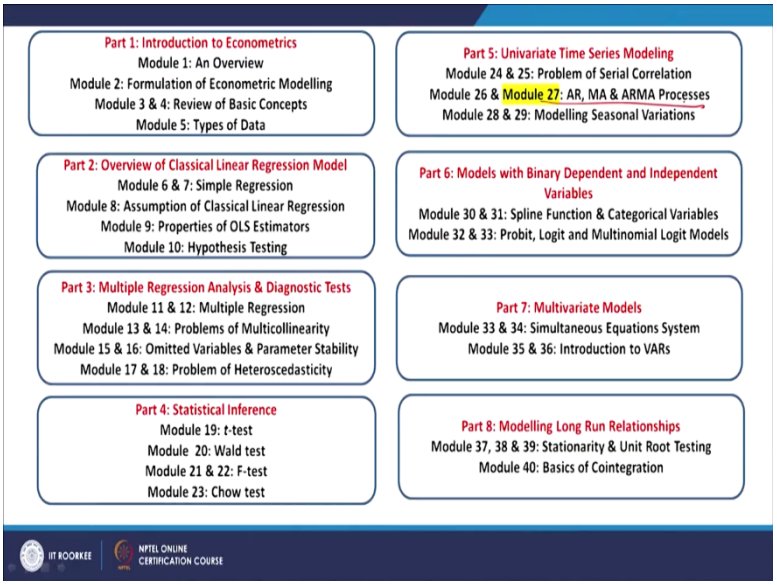
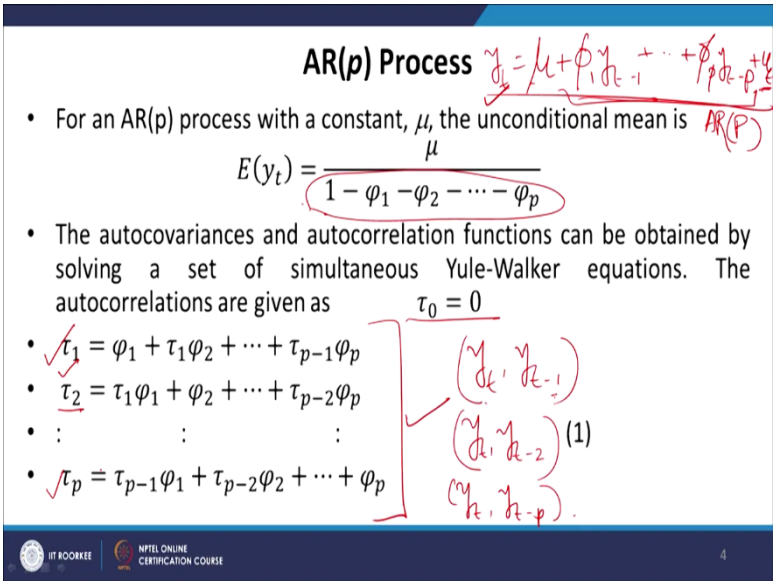
Econometric Modelling
Professor Sujata Kar
Department of Management Studies
Indian Institute of Technology Roorkee
Lecture: 27
AR, MA & ARMA Processes-II

(Refer Slide Time: 00:35)



Hello everyone, this is Module 27 of econometric modelling. In 26, we started discussing AR, MA, and ARMA processes. So, we had discussed there primarily moving average processes that is MA processes. And we also towards the end started discussing AR processes, that is autoregressive processes. Autoregressive processes of order 1 were probably introduced. And here I am going to continue with the discussion on AR processes further along with ARMA processes.

(Refer Slide Time: 01:01)



So, we have discussed already, as I have said the AR (1) process. Now going for a generalization of an AR (p) process with a constant μ, the unconditional mean would be (refer slide time: 1:15).

So, this has 0 mean and constant variance and no autocorrelation or autocovariance between the error terms. So, given this my AR(p) process, we can find out what the conditional mean or the unconditional mean is, so when we go for a calculation of the mean, not conditional upon the values, the previous values.

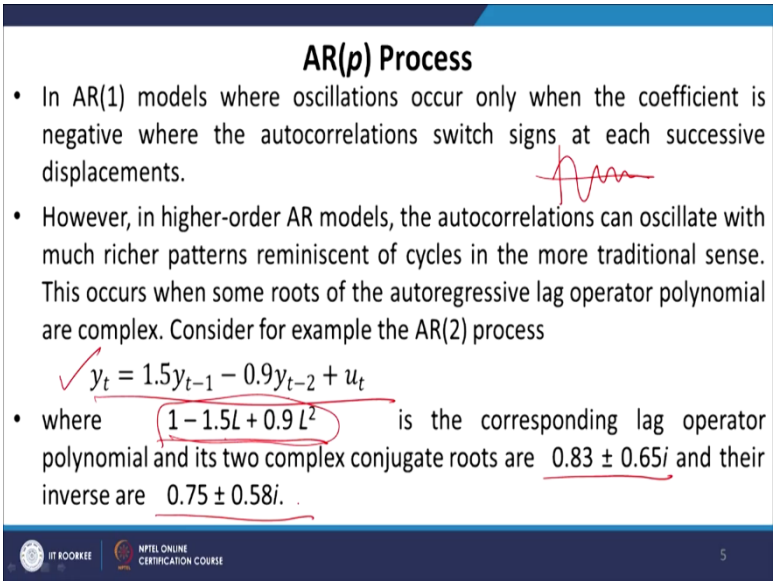
If we condition it upon the previous values then it looks slightly different, but the unconditional mean is, we stake the entire thing all the lagged values to the left-hand side, and this comes to the denominator. (Refer slide: 2:26).

The autocovariance and autocorrelation functions can be obtained by solving a set of simultaneous Yule-Walker equations. Yule-Walker equations were introduced in the context of moving average processes that are recursive. This is a set of recursive equations, so that if we have the initial value, that is, in the context of moving average. if you have the initial autocorrelation value, which is say denoted by t_0, then we can consecutively calculate the rest of the correlations. So, similarly, by applying that Yule-Walker equation we can also calculate the autocorrelations for an AR (p) process, and this could be obtained from this set of

recursive Yule-Walker equations. We call it equation or the set of equations are called equation 1.

So, (refer slide time: 3:34- 4:24)

(Refer Slide Time: 04:23)



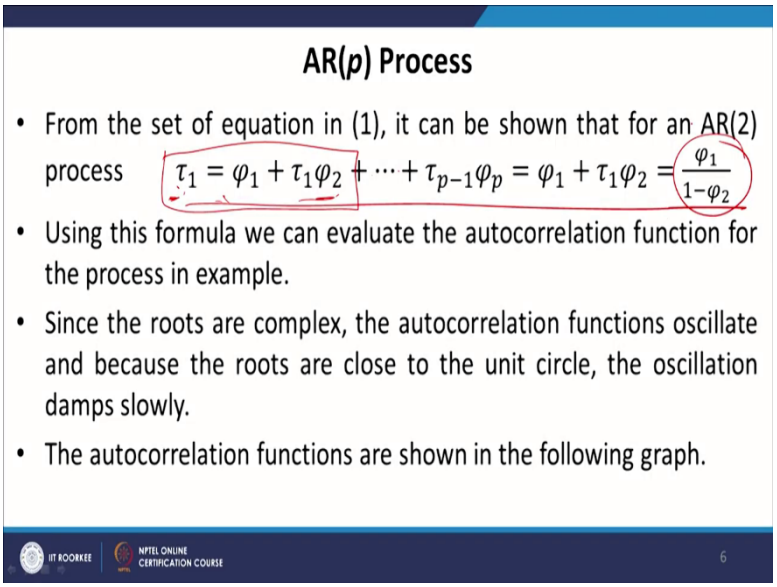
So, in AR(1) models where oscillations occur only when the coefficient is negative where the autocorrelations switch sign at each successive displacement. We have already discussed AR (1) processes, and if you remember, we discussed that, if we have a positive coefficient of the AR(1) term, then we have monotonically decreasing lags and monotonically decreasing autocorrelations.

If we have a negative AR (1) coefficient then the decay is actually through oscillations. So, the decay would be like something like this. However, in higher-order AR models, the autocorrelations can oscillate with much richer patterns reminiscent of cycles in the more traditional sense. So, if we go for higher lags, we can basically replicate the way business cycles are observed.

So, the decay would not be very fast, and we can probably get closer to reality, the way we generally observe the series or time series to move over a period of time. This occurs when some roots of the autoregressive lag operator polynomial are complex. Consider, for example, an AR(2) process, which is like, (Refer slide time: 5:50- 6:34).

So, as you remember that we mentioned the condition for stability or the way we defined covariance stationarity that requests the inverse of the root of the polynomial lag operators to lie inside the unit circle. Alternatively, one can say that the roots of the polynomial lag operator should lie outside the unit circle. So, in case here since, we from the beginning went through the definition of the inverse of the root lying inside the unit circle going by that we must have this inverse of the root lying inside the unit circle.

(Refer Slide Time: 07:14)

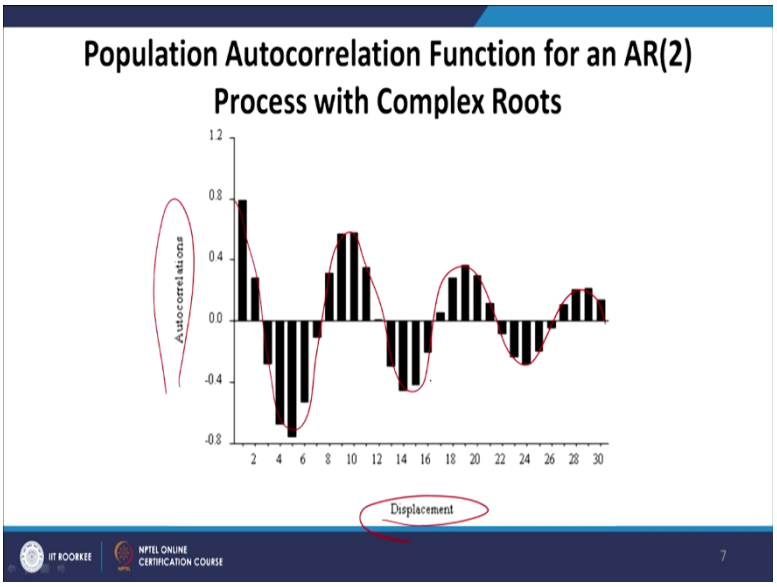


From the set of equations in 1, it can be shown that for an AR (2) process (refer slide time: 7:20- 8:02). So, this is what is the autocorrelation, for an AR (1) process the autocorrelation of the first lag.

Using this formula, we can evaluate the autocorrelation function for the process. In the example, since the roots are complex the autocorrelation functions oscillate and because the roots are close to the unit circle the oscillation damps very slowly. This was also shown earlier that even with real roots if we have the φ or the coefficient very close to 1, then the AR series is much more persistent.

The oscillations are very gradual. And similarly, if the roots are close to 1, then this oscillation would also be very persistent, the decay would be actually very slow. But there will be decay because the complex roots should lie within the unit circle.

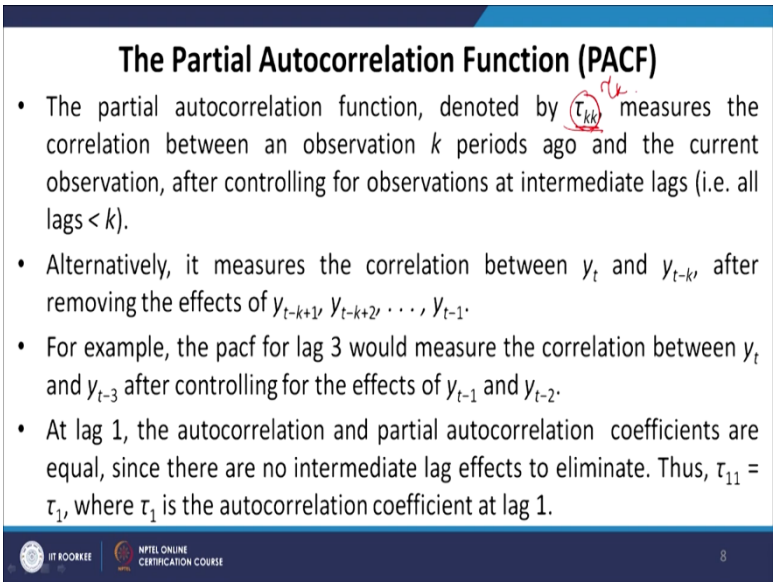
(Refer Slide Time: 09:00)



The autocorrelation functions are shown in the following graph. So, this is a graph, which basically plots the autocorrelation functions, and this is where the displacement. Displacement here refers to lags and these are autocorrelations.

So, this movement is actually closer to the way we could observe the business cycle or some of the economic variables to actually move in real life. Though it is not necessary that the business cycles tend to taper off or they are decaying, but then more persistent oscillations are probably one of the characteristics of many economic series.

(Refer Slide Time: 09:45)

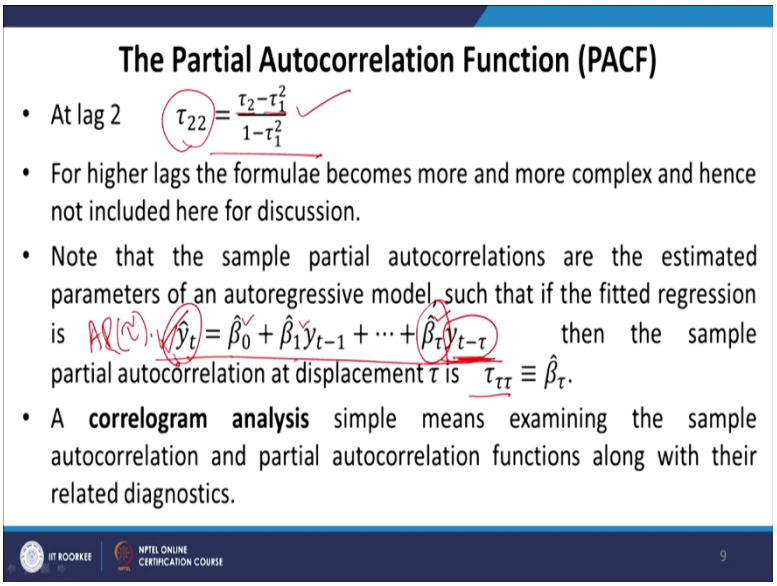


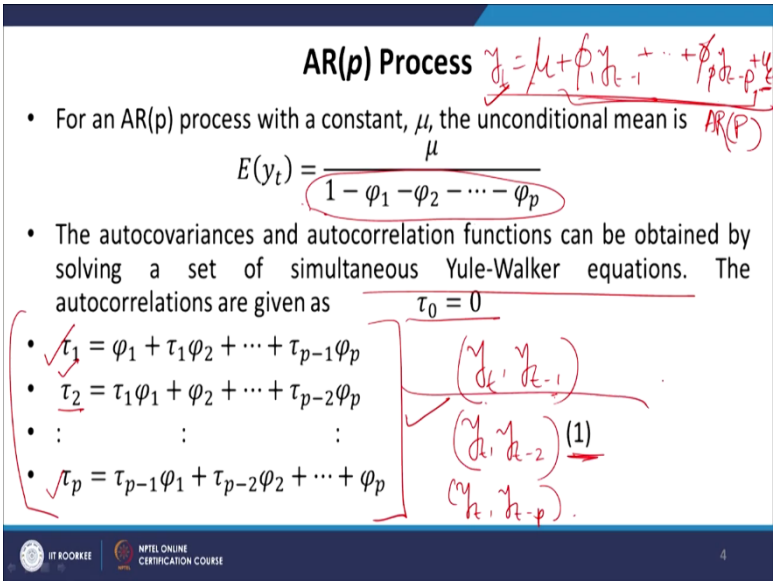
Now, we talk about the partial autocorrelation function. The partial autocorrelation function denoted by τ_{kk} measures the correlation between an observation k period ago and the current observation. So, if we go for a simple autocorrelation this is denoted by τ_k. And since I am considering partial autocorrelation so now, we are having τ_{kk}, as in the subscript. After controlling for observations at intermediate lags that are all lags less than k, we arrive at this partial autocorrelation function.

So, this implies that it directly connects the current observation with k period lagged observation, and the observations in between are not taken into consideration. Alternatively, it measures the correlation between (refer slide time: 10:40) which essentially implies all the intermediate terms have been removed or their impact on the autocorrelation between these two terms are removed.

For example, the PACF for lag 3 would measure the correlation between y_t and y_{t-3}, after controlling for the effects of y_{t-1} and y_{t-2}. At lag 1, the autocorrelation and partial autocorrelation coefficients are equal because there are no intermediate lag effects to eliminate. Thus $\tau_{11} = \tau_1$, where τ_1, is the autocorrelation coefficient at lag 1.

(Refer Slide Time: 11:28)





At lag 2, we have the autocorrelation denoted by , and this is how it is calculated, (refer slide time: 11:40).

Now, how do we arrive at this formula and for higher-order formulas? This is complex, and that is why kept outside the discussion. For higher lags, the formula becomes even more complex and hence not included here for further discussion.

Note that, the sample partial autocorrelations are the estimated parameters of an autoregressive model such that if the fitted regression is (refer slide time: 12:23).

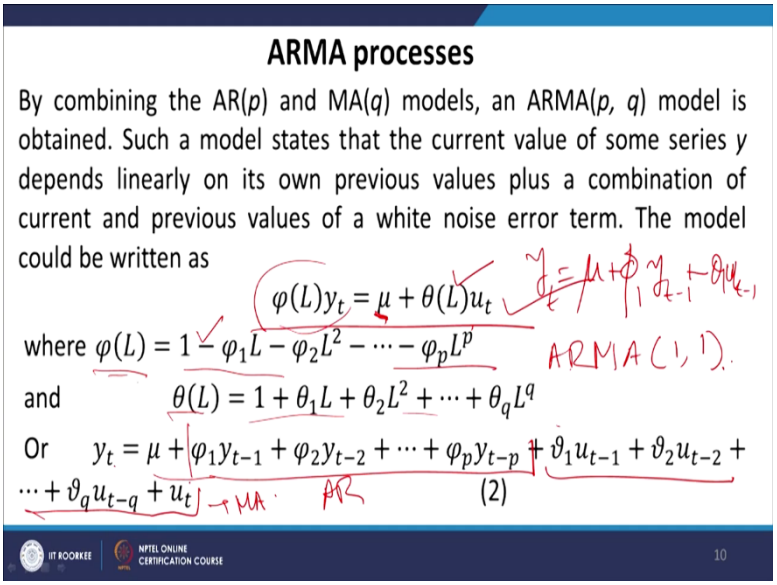
This essentially tells us that, why partial autocorrelation functions are important to us because the partial autocorrelation function gives us exactly the estimated parameter from an autoregression model. So, if I go for an autoregressive model of orders say, τ, so this is an AR (τ) model and this is the estimated part.

So, the initial model would have a (refer slide time: 13:11). A correlogram analysis simply means, examining the sample autocorrelation and partial autocorrelation functions along with their related diagnostics.

So, I showed you equation 1, which consisted of a set of Yule-Walker equations. This is my equation 1, which is a set of simultaneous Yule-Walker equations, we can call it a correlogram which basically helps us in calculating autocorrelation functions. In a similar correlogram, we can also have for partial autocorrelation functions. So, a correlogram

analysis simply means, examining the sample autocorrelation, partial autocorrelation functions along with their related diagnostics.

(Refer Slide Time: 14:52)



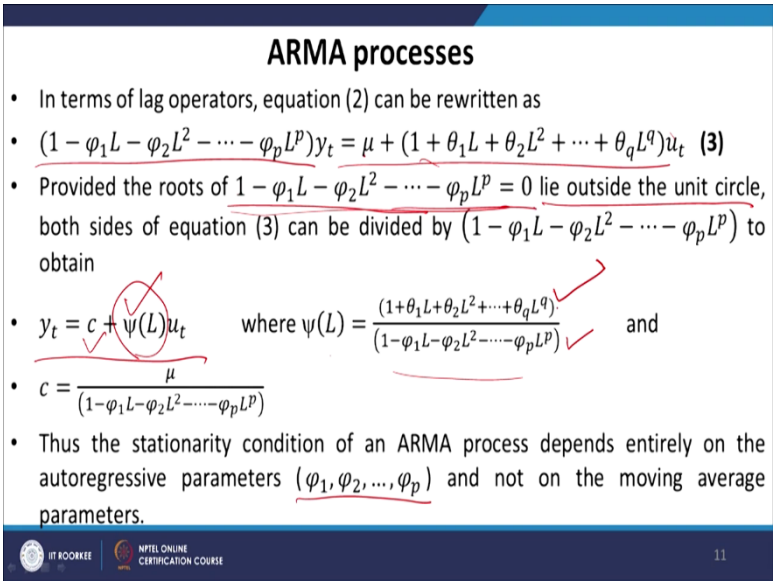
Now, I start talking about ARMA processes. Alternatively, they could be called ARMA processes. By combining the AR (p) and MA (q) models an ARMA(p, q) model is obtained. Such a model states that the current value of some series y depends linearly on its own previous values plus a combination of current and previous values of a white noise error term.

The model could be written as, (refer slide time: 15:22- 15:58), all of them are having this negative sign because they were on the right-hand side, or they're supposed to be on the right-hand side and we bring them to the left-hand side. (Refer slide time: 16:08).

Now, first I wrote it in terms of the lag operator, now I am expanding it. So, what do I have? (refer slide 16:34), μwhich is the constant term plus the lag operator again brought back to the right-hand side, that is, this is the AR components and then we have the MA components. So, these are the MA components.

So, this is an ARMA (p, q) process, p stands for the order of the AR component and q stands for the order of the MA component. So, (refer slide time: 16:59). Similarly, we can go for the ARMA process of order (2, 2), etc.

(Refer Slide Time: 17:23)



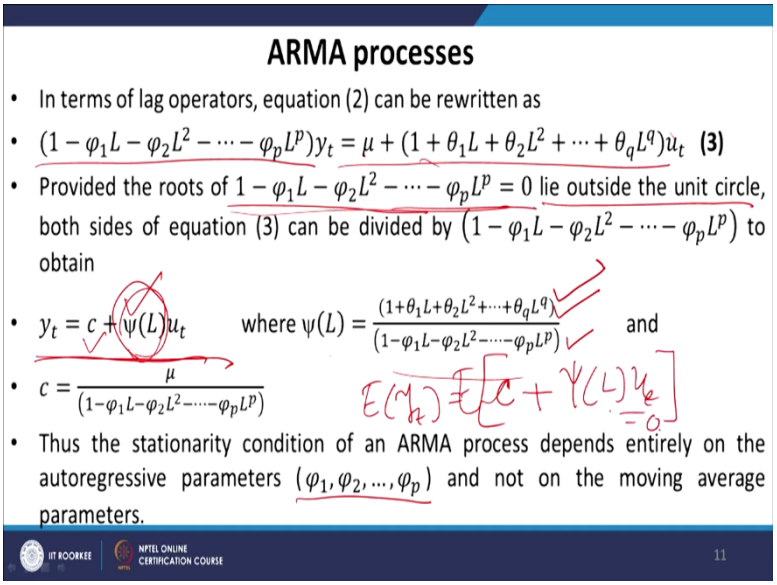
In terms of lag operators, equation 2 can be rewritten like this. So, I am writing equation 3 in order to prove the covariance stationarity condition or show the covariance stationarity condition for an ARMA process. So, I again write it in its previous format, so which we have is (refer slide time: 17:48).

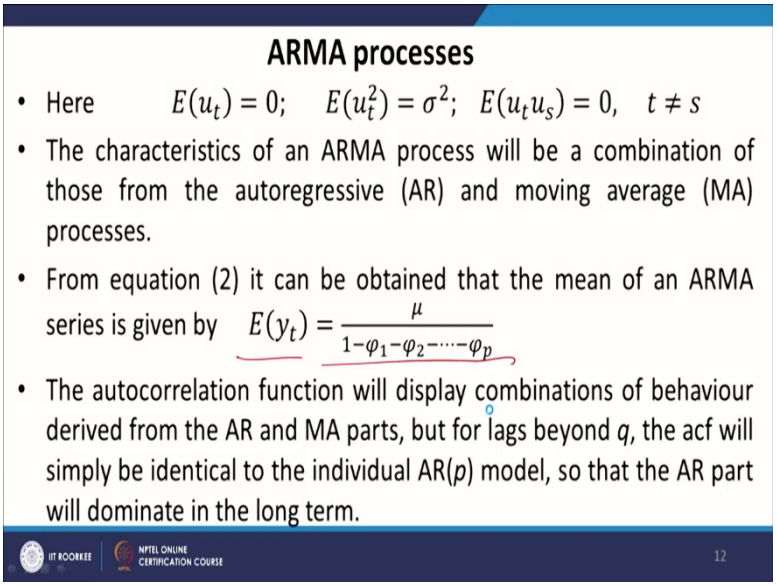
Now, this, we can write provided the roots of (refer slide time: 17:58).

So, the polynomial lag operator of the MA components divided by polynomial lag operators of the AR component gives us $\psi(L)$ and c is the constant term which is μ divided by the polynomial operators of the AR term. Thus, the stationarity condition of an ARMA process depends entirely on the autoregressive parameters, and not on the moving average parameters.

So, this is important to note that, the condition is again, with respect to only the autoregressive parameters. The root of the polynomial lag operator related to or associated with the AR component needs to lie outside the unit circle and we are not concerned about what is happening with the roots of the polynomial lag operator of the MA component.

(Refer Slide Time: 19:24)





So, here again expected value of (refer slide time: 19:27). The characteristic of an ARMA process will be a combination of those from the autoregressive and moving average process.

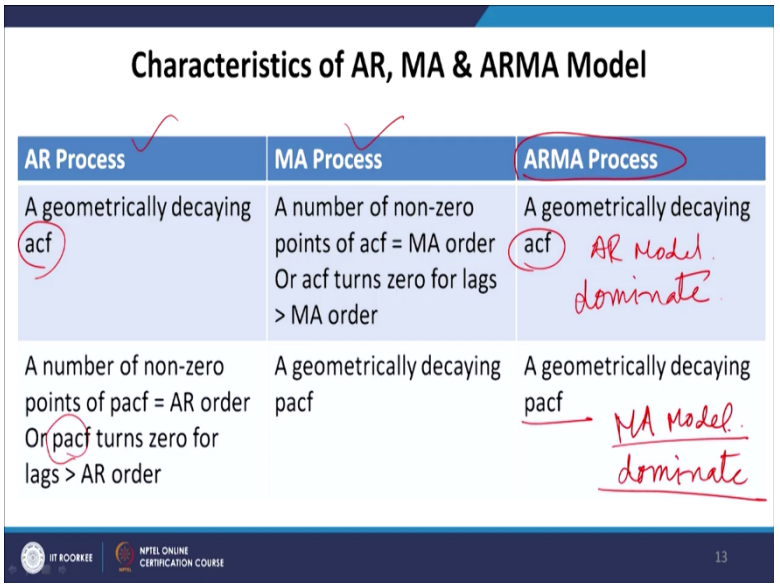
From equation 2, it can be obtained that the mean of an ARMA series is given by expected (refer slide time: 19:27- 20:30).

The autocorrelation function will display a combination of behavior derived from the AR and MA parts, but for lags beyond q, the ACF will simply be identical to the individual AR (p) model, so that the AR part will dominate in the long run.

Because if you remember, when it comes to MA processes, then we have autocorrelations higher than the lag of the MA process all equal to 0. So, for an MA process, the autocorrelations are available or non-zero only up to the order of the process. While that is not the case with AR processes. And because of which autocorrelations will be truncated for MA processes equal to their orders.

As a result of which when you consider ARMA processes, the AR process having longer-term autocorrelations is going to dominate in the long term. And that is why we say that the AR part will dominate in the long term because the MA part is going to have autocorrelations only up to the order of the MA process.

(Refer Slide Time: 21:44)



Now, we discuss the characteristics of AR, MA, and ARMA models. So, this is sort of a comparison between the three models. First, for AR processes, we can experience geometrically decaying ACF or autocorrelation functions. Here ACF refers to autocorrelation functions and PACF refers to partial autocorrelation functions.

So far AR processes a geometrically decaying ACF is observed and for MA processes a number of non-zero points of ACF equal to a MA order alternatively, ACF turns 0 for lags greater than MA order. So, if the MA process has only two orders, then after two lags the ACF is going to be equals to 0, so it gets truncated at lag 2.

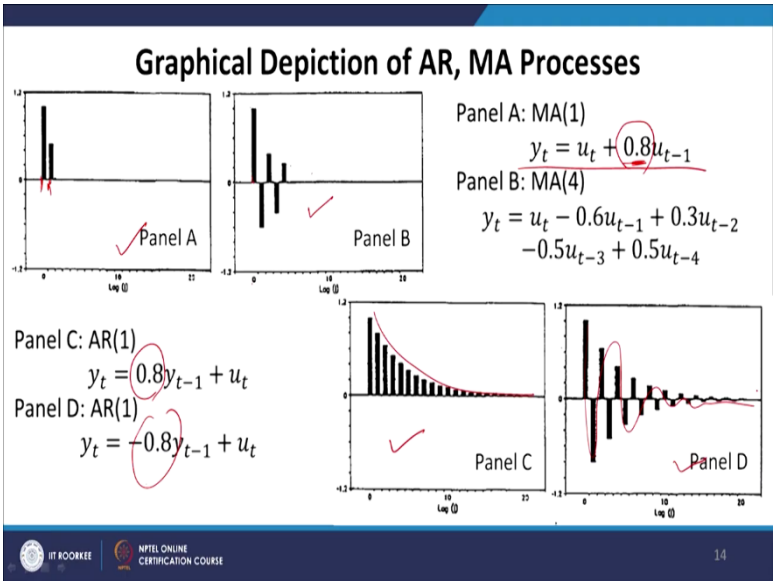
While an AR process even with one lag will be having a geometrically decaying ACF. And for AR, MA process we will have a geometrically decaying ACF because in the case of ACF the AR model will dominate. Now, consider the PACF that is partial autocorrelation functions. A number of nonzero points of PACF equal to AR order or PACF terms 0 for lags greater than AR order.

So, that was a particular peculiar characteristic of AR processes, that for AR process PACF that is partial autocorrelation functions are available only up to the lag of the order. Because, if I am having an AR process of order 3, then I will be having PACF between the current value and the first one between the current value and the second lag, and between the current value and the third lag. After third life there exists nothing, so we would not have any PACF.

But that is not true for MA processes. For the MA process, we have a geometrically decaying PACF. The reason is that if you remember an AR(1) process can actually be expressed as an MA infinite series. And similarly, we can say that a geometrically decaying PACF is observed for an MA process. A geometrically decaying PACF will also be observed for ARMA processes because here the MA model will dominate.

So, as you can see that ARMA process combines both AR processes and MA processes, so it has the dominating characteristics of AR models sometimes, and it has the dominating characteristics of MA models when it is required or when it comes to calculations of PACF.

(Refer Slide Time: 24:33)

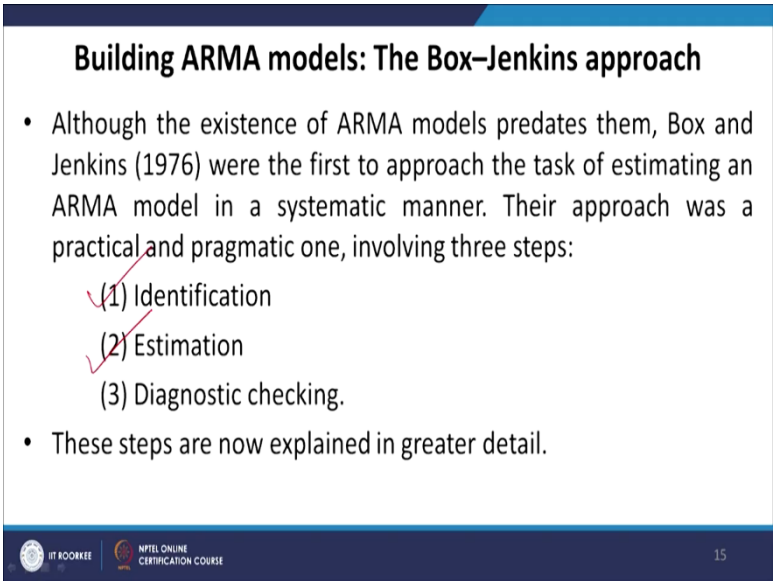


Now I will show a diagram where also these similar diagrams have already been shown earlier, but with some examples, I again show you the graphical depiction of AR and MA processes. So, this is panel A. Panel A has an MA(1) process, where we have a positive coefficient of 1 lag or the first lag. So, this shows that this is the autocorrelation at lag 0, and at lag 1 there is this autocorrelation of 0.8 and after that, it gets truncated.

Now, panel B has an MA (4) process, and again you can see that the first lag is at this is at 0, so this is equal to 1, after that we have -0.6 then +0.3, then - 0.5, and then +0.5, and so on, this gets truncated here. Now, this is panel C. Panel C has an AR (1) process, which has a positive coefficient.

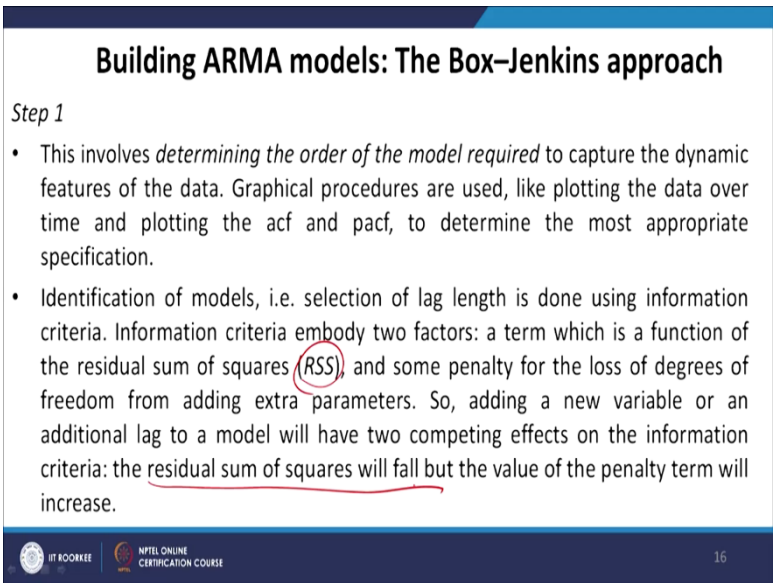
So, the autocorrelation functions actually decay monotonically, and panel D has an autocorrelation function of the AR(1) process with a negative coefficient, so the autocorrelation functions are persistent and it decays with oscillation. So, whatever we have been discussed so far that is again is shown with some examples graphically.

(Refer Slide Time: 26:07)



Now, I talked about one approach suggested by Box-Jenkins for building ARMA models. Although the existence of the ARMA model predicts them, Box-Jenkins was the first to approach the task of estimating an ARMA model in a systematic manner. Their approach was a practical and pragmatic one involving three steps. So, they suggested three steps to follow. The first one is, identification, the second one is, estimation and the third one is diagnostic checking. These steps are now explained in greater detail one by one.

(Refer Slide Time: 26:41)



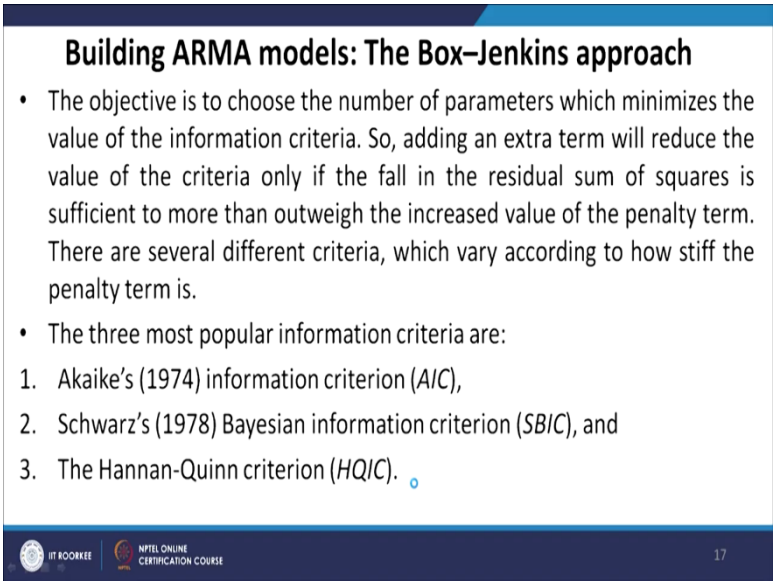
So, talking about step 1. This involves determining the order of the model required to capture the dynamic features of the data. So, we first need to understand how many orders of ARMA processors should be included. Graphical procedures can be used, like plotting the data over time and plotting the ACF and PACF to determine the most appropriate specification.

So, understandably, while considering the AR model, we should go for PACF and while considering the MA model we should go for ACF because that is going to only tell us where the data is going to be truncated, otherwise, for the AR model, ACFs would be monotonically decreasing and for MA models PACF would be monotonically decreasing.

Identification of models, that is selection of lag length is done using information criteria, as well. So, the alternative procedure is to go for certain information criteria. Information criteria embody two factors, a term which is a function of the residual sum of squares, RSS, and some penalty for the loss of degrees of freedom from adding extra parameters.

So, adding a new variable or an additional lag to a model will have two competing effects on the information criteria: the residual sum of square or RSS will fall. We know that with the inclusion of additional variables, ESS or Explained Sum of Square inevitably increase, and RSS inevitably decrease, but the value of the penalty term will increase. So, it will offset the decrease in the residual sum of the square.

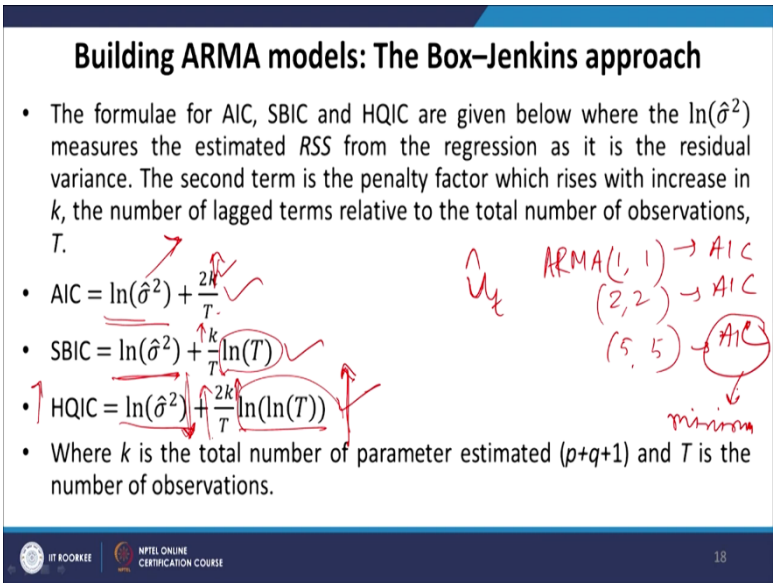
(Refer Slide Time: 28:26)



So, what matters is that the objective is to choose the number of parameters, which minimizes the value of information criteria. So, adding an extra term will reduce the value of the criteria only if the fall in the residual sum of squares is sufficient to more than outweigh the increased value of the penalty term.

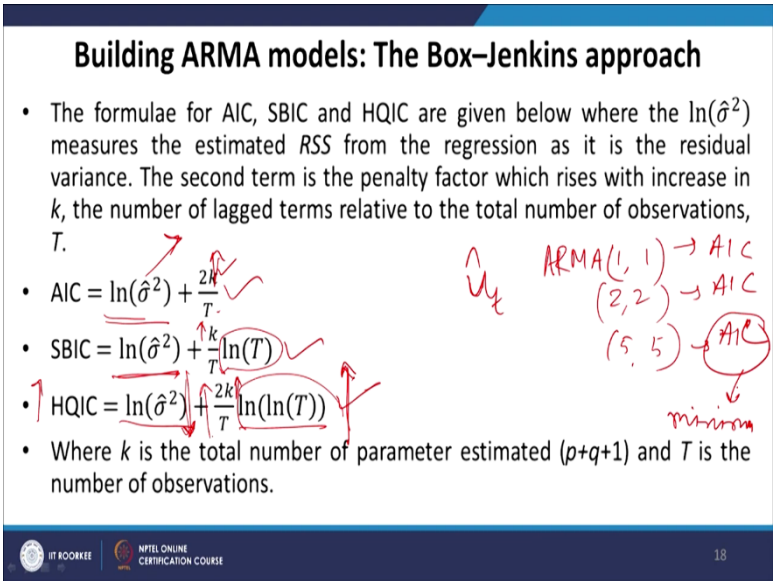
So, if the increase in penalty is offset by a decrease in the RSS, then the overall value of the criteria will decline, then we will try to choose the criteria with the minimum value. There are several different criteria, which vary according to how stiff the penalty term is, so the RSS component remains roughly the same. What varies across different criteria is basically the penalty term.

(Refer Slide Time: 29:20)



Now the three most popular information criteria are Akaike's information criteria that were suggested in 1974. Schwarz's Bayesian information criterion (SBIC), and finally, The Hannan- Quinn criterion (HQIC). These three I am going to discuss here very briefly, that does not mean that other criteria available or suggested in the literature are inferior to this or it is not valid.

(Refer Slide Time: 29:45)



The formula for AIC, SBIC, and HQIC are given below, where (refer slide time: 29:50-30:17).

The second term is the penalty term and that varies across criteria. So, this rises with increasing k, the number of lag terms relative to the total number of observations T. So, the moment we have more and more lags included, which simply implies that the number of independent variables increases, so k is going to go up.

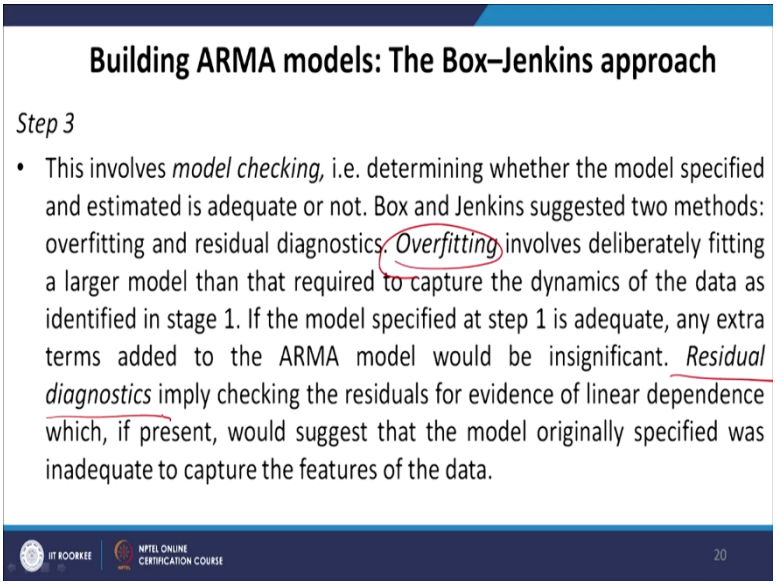
The numerator is going up, the denominator remains constant. Her, also the denominator and a part of the numerator remains constant, but a part of the numerator increases. So, these terms are expected to increase with an increase in the number of lags included in a regression. Now, if this acts as a penalty term because this is going to decrease inevitably.

So, if this fall is outweighed by this increase, then overall, the criteria value increases otherwise it decreases. We should go for the minimum value of the criteria that is by comparing alternative models we should go for the value of the criteria which gives us the minimum value.

So, for example, I estimate ARMA models with (1, 1) that is 1 lag, then 2 lags, then 5 lags and then compute AIC values for each model or I can go for also SBIC or HQIC any criteria I consider. And I will basically choose that AIC value which is the minimum between all these models. So, that is how the criterion is or the criteria are used. Now, here k is the total

number of parameters estimated or variables considered lag lengths considered, so this equals to $p + q + 1$ and T is the number of observations.

(Refer Slide Time: 32:09)



Between the criteria mentioned SBIC is strongly consistent, but inefficient, and AIC is not consistent but is generally more efficient. What does it mean? This means that SBIC will asymptotically deliver the correct model order, that is, if we work with the large sample then SBIC is more consistent in the sense that it gives us a more correct model or correct model order.

While AIC will deliver on average two larger models specifically relative to SBIC, even with an infinite amount of data. So, this has no asymptotic property, regardless of whether you are working with the small data, or a large data AIC is going to deliver a relatively larger model compared to SBIC.

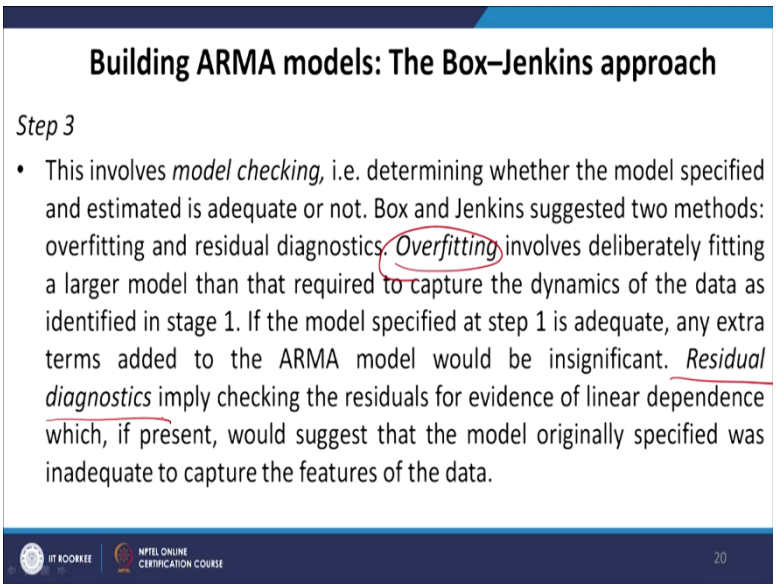
On the other hand, the average variation in selected model orders from different samples within a given population will be greater in the context of SBIC than AIC because we have mentioned that AIC is more efficient. So, the variations in the model suggested by AIC are less when we repeatedly go for samples from the same population, while the variations in SBIC would be high.

So, if we are working with a large sample, then it is always preferable to go for SBIC. Overall then no criteria are definitely superior to others, all of them are having disadvantages

and advantages. And we have compared the two extreme cases SBIC and AIC, HQIC lies in between.

Then talking about step 2, this involves estimation of the parameters of the model specified in step 1. This can be done using least squares, so another technique known as maximum likelihood depending on the model.

(Refer Slide Time: 33:56)



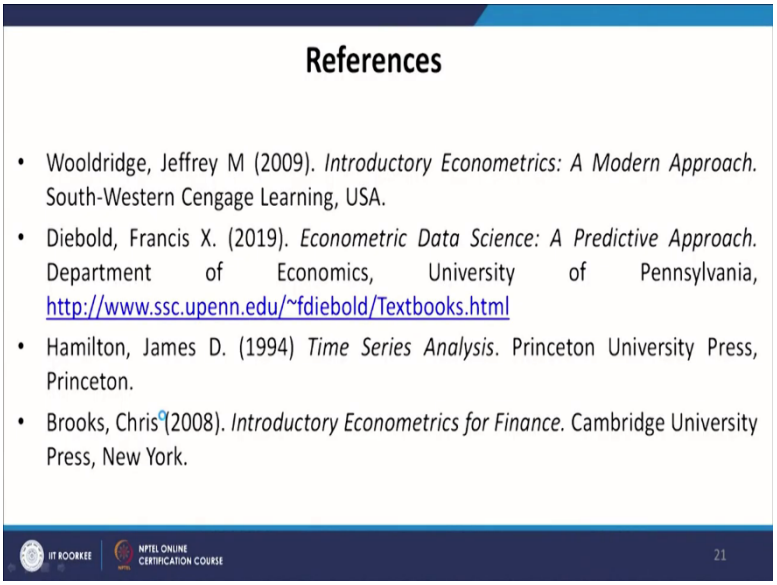
In step 3, we go for model checking, that is, determining whether the model-specified and estimated is adequate or not. Box-Jenkins suggested two methods. First, overfitting and then residual diagnostics. Overfitting involves deliberately fitting a larger model than that required to capture the dynamics of the data, as identified in stage 1.

If the model specified at step 1 is adequate, any extra terms added to the ARMA model would be insignificant. On the other hand, residual diagnostics imply checking the residuals for evidence of linear dependence, which if present, would suggest that the model originally specified was inadequate to capture the features of the data.

So, residual diagnostics check for the characteristics of the residuals, and if the residuals are well behaved then we go ahead with estimation or we think that the model is adequate otherwise, it is not. ARMA models as suggested by Box-Jenkins' approach actually constitute a large part of a statistical time series analysis where very briefly just presented a sketch here

just to give you some idea about it. It is not even closer to the actual analysis that is suggested and presented by Box-Jenkins.

(Refer Slide Time: 35:20)



These are the references that I had followed. And this completes the discussion on AR, MA, and ARMA models. Thank you.