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

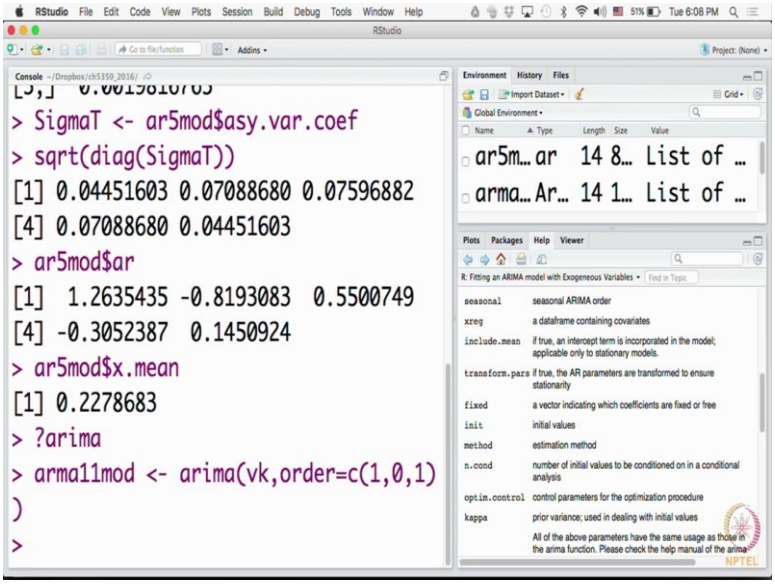
Lecture – 52
Lecture 22C - Models for Linear Stationary Processes with R Demonstrations16

Fortunately, you have ARIMA routine which is very generic routine which can be used for estimating AR models, MA models, ARMA models, ARIMA models, even seasonal ARIMA models and many you know different types of models. Then why did not I start off with that? As I said you should know that there exists specialize routines for estimating AR models which use different set of algorithms they AR model, sorry ARIMA routine in our uses maximum like would estimating which are computationally intensive. So, if you have large data set you may not want to really straight away the use ARIMA and MLEs are known to be notorious for that kind of notorious when it comes to computational complexities and so on.

So, it may be good idea if you know that you are using AR models to stick to your AR routine and get you work done because two advantages: one the algorithms that are used in this are fairly simple Yule Walker (Refer Time: 01:16) square so on. You have computationally efficient ways of calculating them. And also that it gives you unique estimates, whereas ARIMA uses a MLE and any maximum like would estimation algorithm 99.9 percent of them are all non-linear lease square problem, and therefore you will be given only local optimum it I think uses some kind of a Gaussian Newton algorithm or Fisher's scoring algorithm.

So, the point is that you get you unique estimates and algorithms are simple in AR.

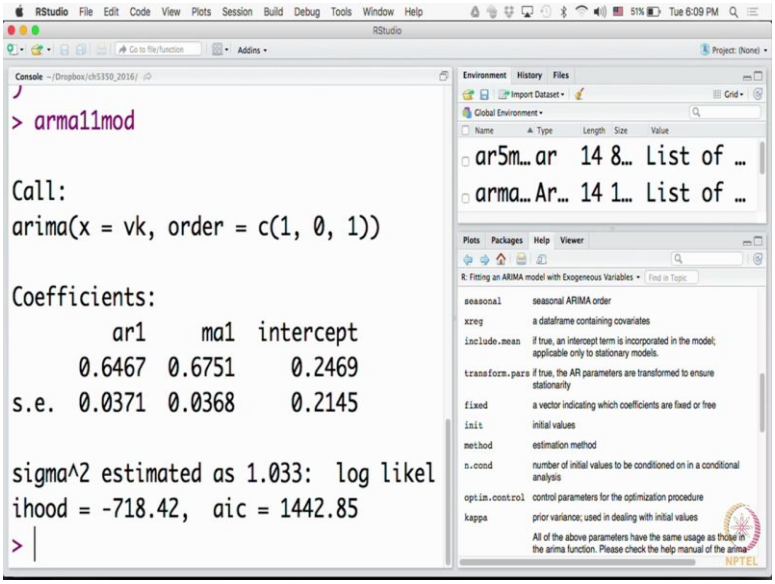
(Refer Slide Time: 01:51)



So, very quickly let us look at the ARIMA, here again of course it asking which one will stick to this, because the ARIMA that comes to TSA as some advantages I am not going discuss that at the moment it suffices to know again we have supply for example the series of course and the orders. And then again specify d mean. And there is one option that I would like to draw your attention to which is transform dot pa pars. This transform dot parameters essentially I do not know how will you can read it says if true, that means it is a logical variable; the AR parameters are transformed to give you stationarity and I have to go and figure out as to what it does to give you stationarity there. Essentially this forcing some kind of stability on the model, it pulls back if it finds AR model in the outside the unit circle it kind of pulls back somehow and forces it, but let us not worry about it now but I just wanted to point that out.

So, let us quickly fit an ARMA 1 1 model and I am going to fit an ARMA as I said 1 1 because I do not have any information and I have fit it. And d mean has said to be true, that means it will also estimate either the alpha or the mu. Now the nice thing about using this routine is the way unfortunately it is not these are things are not structured very well because it is an open source software of whatever I hope are developers are hearing us, but the way the ARIMA model is reported is a lot better than the way and AR model is reported. And that is why even though it is using a maximum likelihood algorithm I would prefer this for example, it do the diagnosis for this.

(Refer Slide Time: 03:42)



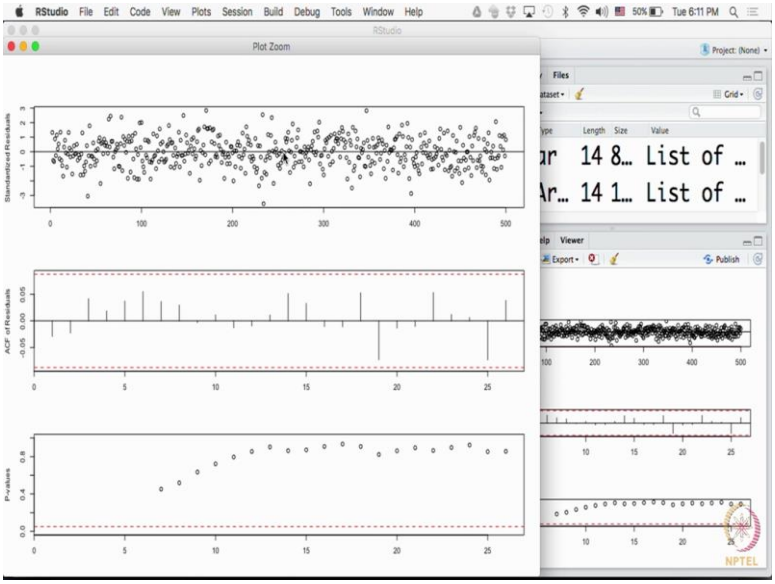
So, for the example here I can ask for the model and it straight away calculates standard errors for me, because then I can use this valuable time to post another massage on Facebook. Look at how much time I wasted in computing all the standard errors and so on. And because that is the object at the end of the day, maximize time on social media. So, anyway you had a question.

Student: (Refer Time: 04:15).

Sorry, so they order 1 0 1 there it is a very secrete thing. So, you should be able to guess now. Those are the orders for the AR the integrating effect if there is 1 and then an MA order; anyway, thanks for brining that all. So, we since we are fitting ARMA 1 1 we are saying that and there is not integrating effect I believe in the model in the series sorry, and the orders are values are self explanted. So, if you are fitting ARMA 2 1 you would be supplying 2 coma 1 coma 1. So, the see there is a allowing you to pass on a vector because you know c and r actually works gives your vector.

Let us quickly look at this model here; it reports to estimates and it also gives me intersect term and then it also gives me standard errors, but we already being told by some one that we should not look at the standard errors until and unless we are convinced that the residual survive. And we do not know who said it, but someone said it. So, let us actually go back and look at the residuals.

(Refer Slide Time: 05:28)



Now, there is exists one routine called tsdiag, ts basically diagnostics for your time series model and you can pass on this model object that you have estimated to that and it generates the required diagnostics for you. So for example here, I have estimated ARMA 1 1 mod and I want to do a diagnosis of that model. So, it generates its plot let me quickly explain what this plot is showing you. The top plot is showing what is known as the standardized residuals. So, it is a standardized with respect to something, do not worry just showing residuals for you in a standardized form.

Now obviously by looking at the residuals is a exactly I got my model you cannot say that, you cannot jump with joy just like you have to look at the second plot which is showing you the ACF. What is ACF telling me? Are the residuals do they pass the whiteness stress or not? No [FL] where?

Student: (Refer Time: 06:30).

Sorry.

Student: (Refer Time: 06:33)

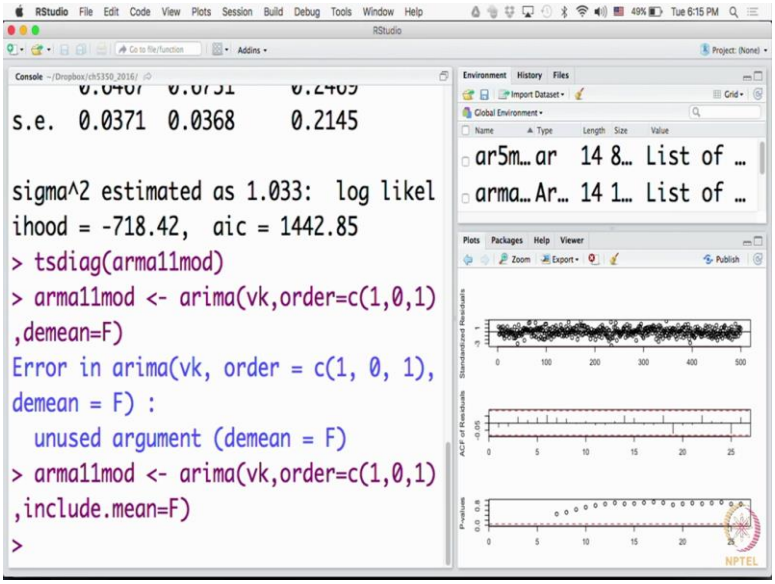
No there are dash lines, are you able to see the red dash lines there? You are able to see? Maybe if you are gone for shopping or cloth store you would be able to see all of them because the plot may be (Refer Time: 06:49) visible a very very careful when we go for shopping. So, there are these red lines which are the significant bands and all the

estimates are there within that significant band. And that is a way of; we will understand truly what is the meaning of the significants bands later on, but as of now the way to use them is if the all the estimates are within the band then we conclude that none of the ACF estimates at non zero lags is significant.

Then there is a bottom plot which has got to do with what I said earlier (Refer Time: 07:22), but I am going to skip that for now when we talk about the (Refer Time: 07:28) you will understand what the plot is generating. But that is also telling me that the model is actually you know residual are white, its passes the whiteness the residual pass the whiteness test. So, the conclusion is now this ARMA 1 1 model is satisfactory with respect to residual analysis. So, the next check up is whether it has over written or is been over trained; and that is why we are look at the standard errors. And clearly it tell us that the parameter estimates namely the d 1 and c 1 has r significant, because the errors are small you can I mean you can construct the confidence region.

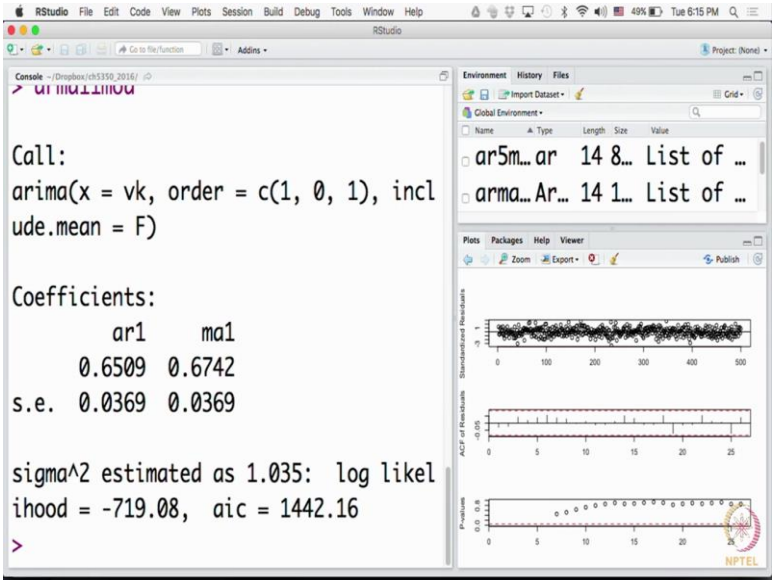
But it is say about intercept term? They estimate is given 0.2469 and the error is also given 0.2145, error is as big almost has estimate. So, what does it tell us about if I were to look at this kind of a hypothesis this is an mu or alpha? Say is that the; (Refer Time: 08:37) hold I mean I fail to reject the null hypothesis, because again you apply the same principle you construct the confidence region it should not contain a 0. But it does, do not tell me that it is, it definitely contain there is nothing definitely it contain or not that is all beyond that it is all only intensity of how you say it. But unfortunately there is no accounting for that in statistic you can say it very lightly in a sleepy manner it does not matter it says truth is truth.

(Refer Slide Time: 09:23)



So, the point is now strictly speaking I should go back and also now undo this and say I do not want to estimate the mean. So, we should do that and that is the most important step. So, we say do not be mean; sorry what happen demean, I think may be include dot mean, so falls.

(Refer Slide Time: 09:58)



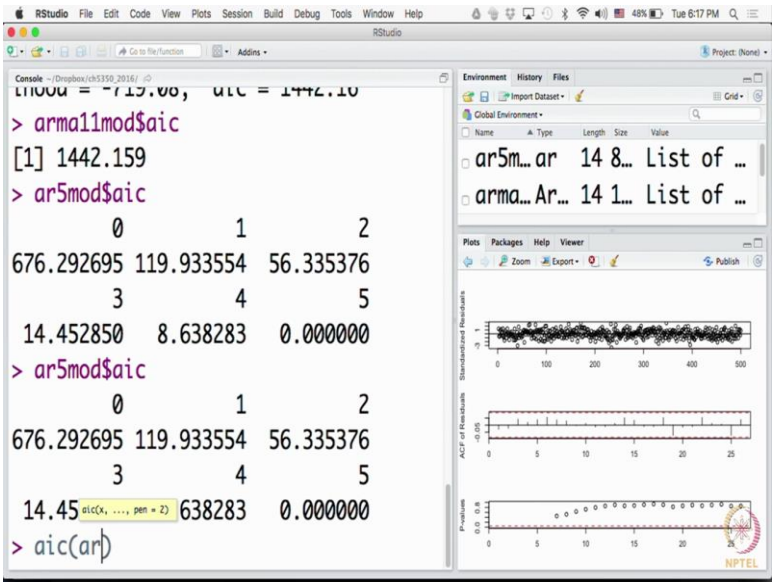
So, in this unfortunately it is not demean it is include dot mean so you can see that there are different thing. And now when you ask for the ARMA 1 1 model, if you ask me not to estimate I estimate it.

So, now this is satisfactory of course, the test for model is completed only when we use a cross validation set and do the prediction, but we will not going to that at the moment. So, the last thing that you want to answer now what is that?

Student: (Refer Time: 10:19).

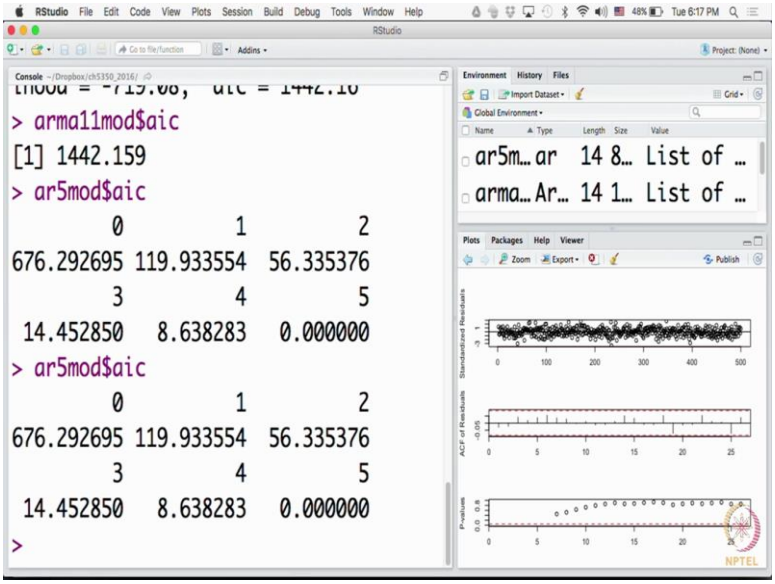
So, I have this model of the cell phone the other model of cell phone both seem to be doing the same thing there are some considerations and so on which one should I pick. So as I said one of the ways, please remember that one of the ways for that is or the method that is used in model selection, model selection still remains one of the hot topics in modeling not just time series modeling everywhere as to which model to pick even in fashion that is that is an issue, anyway. So seriously, models can become out dated also; anyway.

(Refer Slide Time: 11:09)



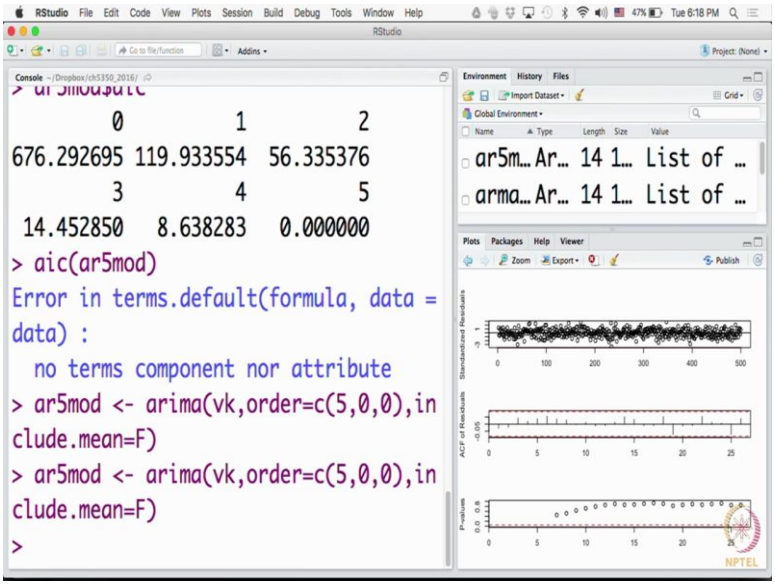
So, let us actually ask for, in fact I do not have to supply aic here, I can ask for the Akaike information criterion. What is the value that has come about for this model and it automatically computes that as a part of model estimation, this is the value that I get. This absolute value does not mean much to me, what matters is which has lower aic that is the one and you pick the model that has a lower aic. I am not explaining what is aic at the moment, but you pick the one that has the lower aic; sorry something is happen.

(Refer Slide Time: 11:46)



Unfortunately, I think there is a problem with aic looking to it, but ideally it should give me the aic calculate in the same way.

(Refer Slide Time: 12:03)



So, what we shall do is, in fact I understand what the issue is; let us actually use the same ARIMA routine to estimate the AR model as well. So, that things are now consistent everything whatever methods are being the used are the same, so we just go through this additional step we say order is c 5, 0, 0, because we want to estimate only an AR model

also we will ask the mean estimation to be taken out; sorry something else. What are the attribute? Is something wrong? Do you notice any error?

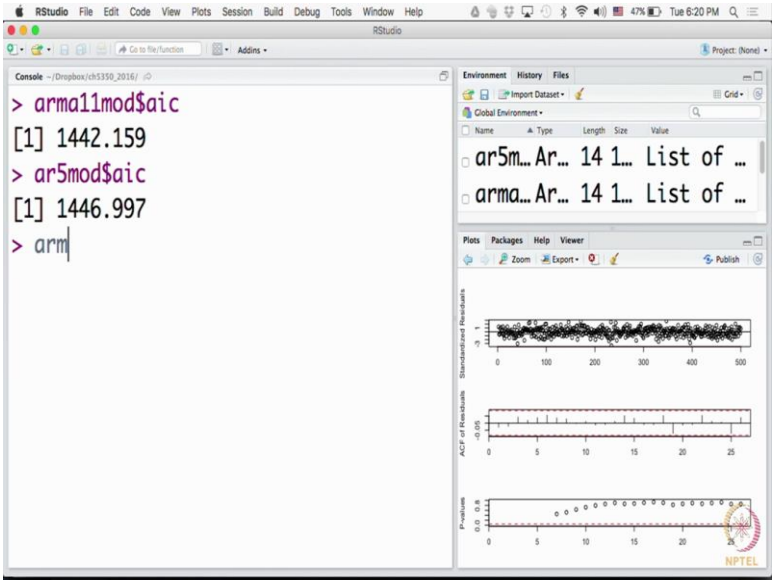
Student: (Refer Time: 12:58).

Then earlier, what was the problem.

Student: (Refer Time: 13:01).

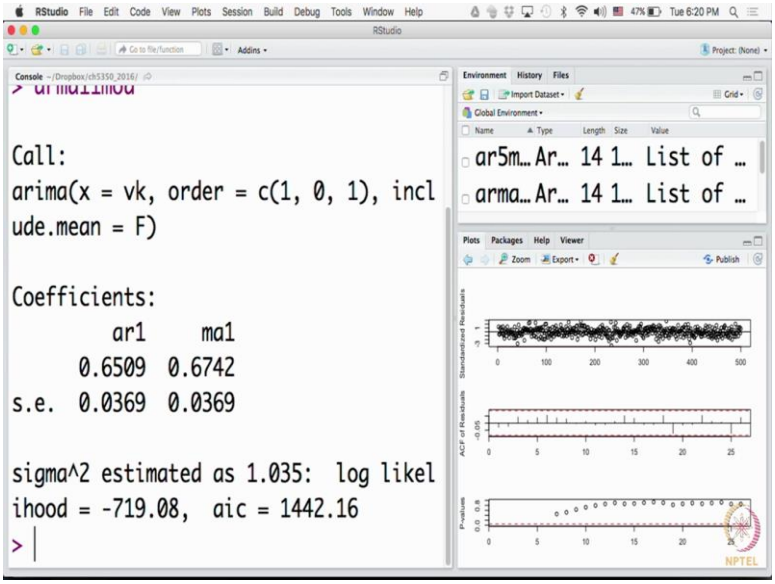
Sorry, I think I need to stop them; really sorry thank for telling me that, so error my side. So, now we can ask for the aic, so let us clear yes there is no error. This is aic for the ARMA 1 1 and this is the aic for AR 5. And as I said you pick the one that is actually has a lower aic that is a theory. What the aic dose is? It actually looks at the trade of between the model complexities; that is how many parameters you have been included in your model versus how well it has explained the series in terms of minimizing the some square prediction error.

(Refer Slide Time: 13:21)



So, the AR 5 has more parameters. Yes, so it may have done may be some good job of predicting the series. ARMA 1 1 as only two parameters in it, it also succeeded in doing giving a good fit your series.

(Refer Slide Time: 14:19)



But when you look at trade of it says it this is better, ARMA 1 1 is better. So, our winner I do not have a sound for it, but our winner is this.

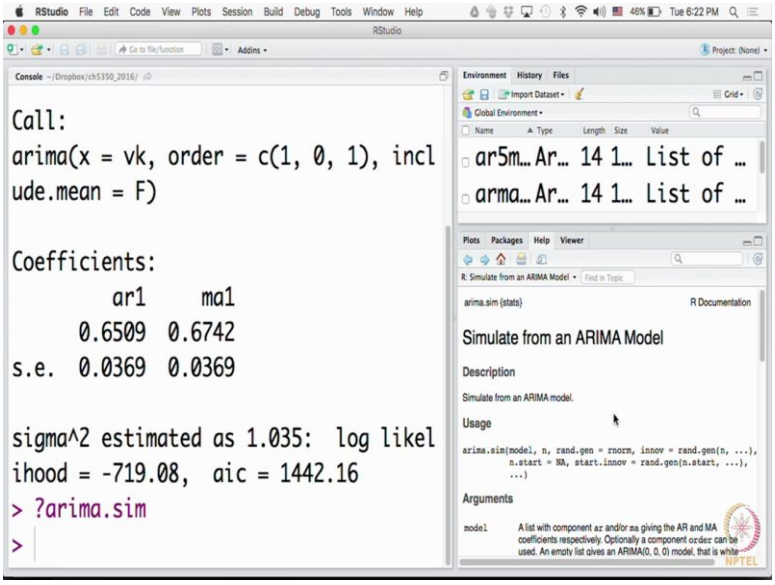
Now time to reveal what was the process that I have used to simulate this; so let us do it here. So, this is the model that I have used, I have used an ARMA 1 1. So, do not tell me a back I mean nowhere in our systemic procedure we have use that knowledge. If you are given me this series and if you are made me blind to the weight was invited we would still follow the same procedure. But by following a systemic procedure at least in theory we have shown if the underline process is ARMA we are manage to pick that. In reality the processes are going to be more than ARMA, you have ARMA as you have so many things, so then on going to be ARMAs for you.

But, this kind of a conviction is necessary for anyone to demonstrate that- yes this procedure actually does pick the right model when the truth is that. Again that is the standard procedure that used everywhere. So, also this way to simulate and ARIMA series you can see as specify the model as a list I specify the coefficients, the only question is when we specify coefficient it is asking you to express give it minus d 1 minus d 2 and so on be careful.

So, it expresses you to give the AR coefficients when you write it as v k equals something. The MA coefficients are same and I have said order is c 1 0 1 which means I do not want to have any integrating effect, and n is simply the number of observations.

There are other options in this ARIMA dot sim; for example, here it automatically generates the e k internally generates the e k through r norm and then solves the difference equation.

(Refer Slide Time: 16:27)



You can externally supply you can say there is an option here when you pull up the help on arima dot sim, I do not know how well you can see but it shows here ran dot gan; you can supply by default r norm you can use uniform white noise any other white noise if you want. Or else you can supply a function for example, to generate the innovations. And then there is something called n dot start.

Now, if you go down the help it could say n dot start is nothing but the burn in period. Comes in to picture, remember we have shown and you have seen that the as AR process is not stationary per se it is only asymptotically stationary, if you let it settle down then it become stationary. So, what do you mean by let it settle down, how long should you allowed to settle down?

That is called the burn in period and that is what this n dot start is. One does not have to specify fortunately it automatically kind of figures out how long it takes; although we are asking for 500 observations. Therefore it is generating more than 500 observations and throwing away the first and giving the rest to you so that you are working with the stationary series. That is that something that one can keep in mind.

Now that is 6 o’clock we will wind up.