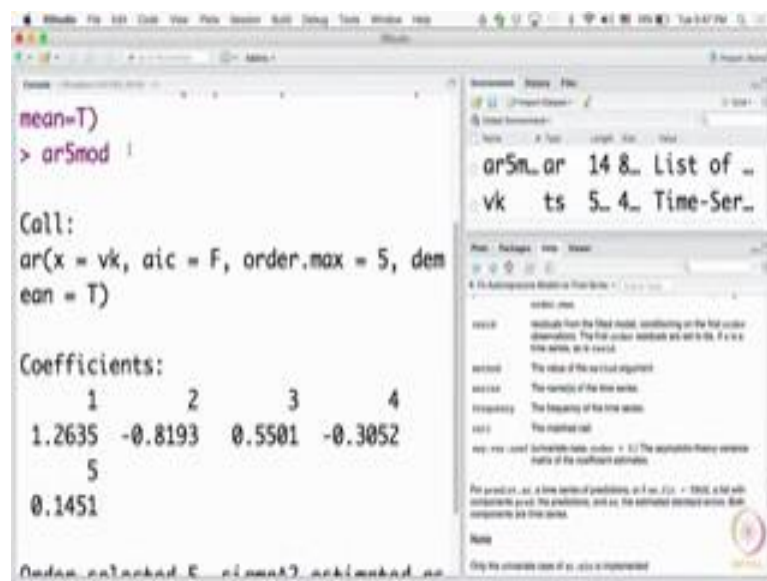**Applied Time-Series Analysis**
Prof. Arun. K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture – 51
Lecture 22B - Models for Linear Stationary Processes with R Demonstrations 15

This is the kind of learning that one has to do before using a routine, do not just jump and use it.

(Refer Slide Time: 00:20)

Let us say you are going to fit an AR model of 5th order and of course, R studio is nice tab, will pull up the arguments, I do not have to say x if I am specifying in the right order I can just specify the series and then I need to specify aic, what we shall do is we shall specify aic as falls; that means, we will freeze the order to our guess because I do not want to pamper you guys and girls too, to using aic all the time. So, that you make some responsible decisions. So, now, what this routine will do is we will fit a 5th order AR that is how this is not going to do any search for me and demean by default is set to true. So, that is it I have estimated the model now to I can actually ask for the model these are the coefficients I have just typed in the name of the object in the model object and it shows what is the call that I have used just in case may be I was typing in with one eye closed its telling me what exactly you typed do not blame me.

Just a reminder that you have asked me to do this, I have done this and it is a good check, many a time, it is a good check because sometimes if even if you do not have the history of commands it is stored whatever the call that we have used to stored in the object and that is a good practice programming practice and so the coefficients are given that is estimates of d1 to d5 and the order selected anyways 5 we have told that and the sigma square e is also estimated for you, 1.061 at this stage I have no idea whether this is correct or wrong, in correct this since it is good model or not.

What is the first check for us look at the residuals aic for the residuals whether there white in color or not. So, we can ask for that, alright. So, this is what we are going to generate oops sorry. So, now, this is something that these are the things that you need to know because generally what happens is if the assignment deadline is midnight then you start working at 11 o'clock, you think everything is going to be just a smooth game, I know exactly the routines that I am going to use and there you will through a error at 11:45. Then you frantically running on what is this n a dot fail, looks like I am going to fail by in submitting this and so on.

What do you think? Can you any of you really guess what is the source of this error that is throwing up?

Student: Sir missing value.

Ha, why there should be missing value? Who went missing? What went missing? What is the missing?

Student: The type of geometric series. So, that is it is not at time series.

No, no, it does not say that it clearly says error in any dot fail right, missing values in the object; it does not say faulty object or incompatible object, nothing like that.

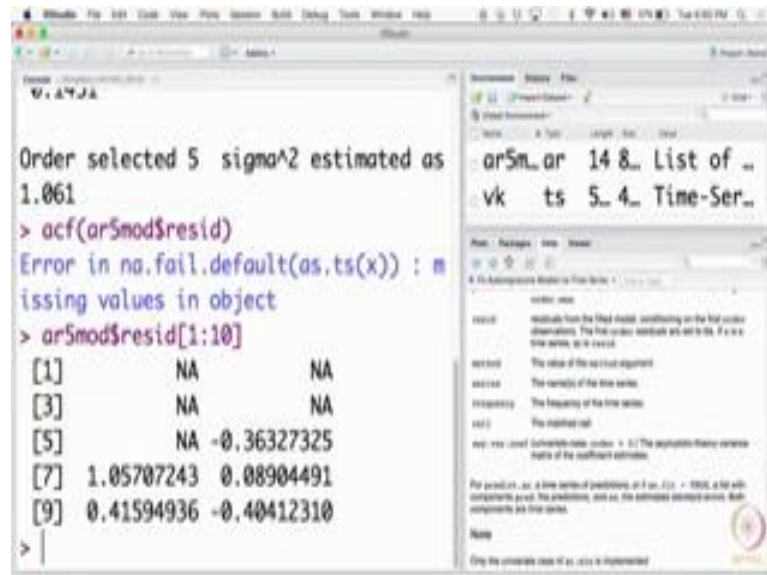Student: Sir (Refer time: 03:39) gone away.

Sorry

Student: Residual point (Refer Time: 03:41).

What has happened? So, the first thing that one because it is telling that there are missing values lets open up the residuals and may be look at the first few observation and see if

indeed there is a missing observation then ask if there is why. So, let us ask here for the first 10 observations let us say.

(Refer Slide Time: 04:14)

There are quite a few missing values right how many do we have?

Student: (Refer Time: 04:18).

5, why?

Student: (Refer Time: 04:21).

Sorry, why they are not available?

Student: (Refer Time: 04:28).

This is the story. So, that is why one should know first the theory and then also the practice, these are practical aspects, the first 5 are missing it clearly says I do not know because you have not predicted, your prediction began only from k equals 12 in this case accounting begins from 1. So, it says you have started to predict only from 6 onwards.

Therefore what should we do? We should just ignore those 5 and let us see if I can specify that or we can a really start our residuals itself from 6 onwards, let us see if I can if acf actually r here in oops sorry if it says omit, does it take this.

(Refer Slide Time: 05:30)
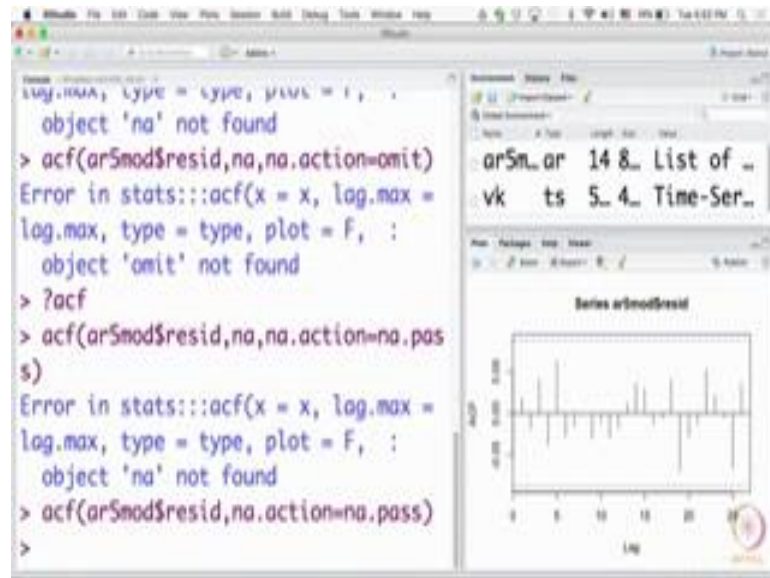
Now, we can go back to acf and ask what is it correct one. So, at the moment what I have done is actually installed the package called TSA and I have told about this as before as well there is a package called time series analysis which also comes with this acf routine and straightly embellished and that is why when I ask for help on acf it shows there are 2 acfs, which one, which [FL] do you want, the stats [FL] or the TSA [FL]. Now since I have loaded TSA after the stats has been loaded, it uses the most recent one from the TSA and let us ask what is na action how to handle missing data. So, apparently it says pass which means let not the security guy give you a trouble even then it says.

Student: na.

I am sorry, yeah, twice that is the reason I think fine that is it. So, the security guy who is sitting their screening for na has been by passed he has been allow to give a pass and therefore, it has omit they ignore those n a values and this given you this plot and you can see here this acf plot actually begins from lag one unlike your regular acf plot because if I do not know how well you can see the help there is an option called drop dot lag dot 0 it is a logical 1 and they default is true if you set it to false, you will get your traditional plots that we have been using. So, drop dot lag dot 0 equals true, sorry, false and then you get this form plot.
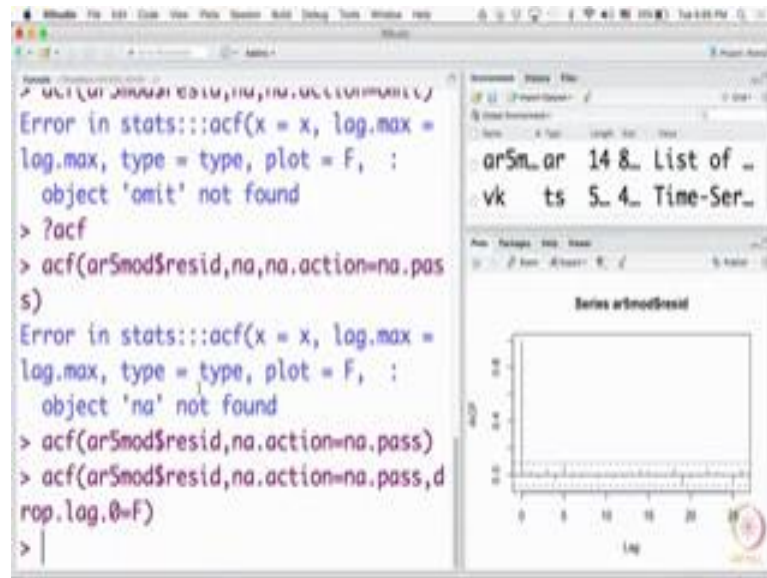
(Refer Slide Time: 07:22)
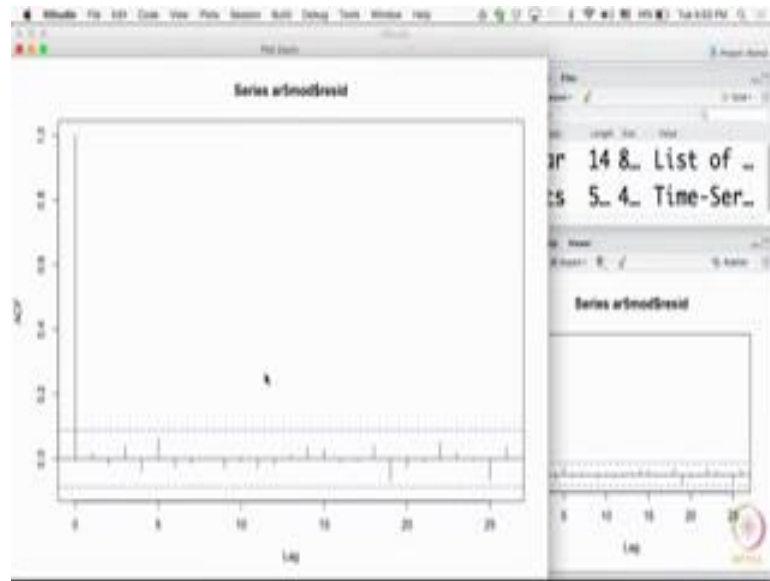
This is the acf of the residuals at we have estimated what do you think are we kind of convinced right that the residuals are white.

We will learn a more formal test of whiteness later on known as Box Ljung Pierce test which is better than this which as a better power in term in terms of hypothesis testing terminology this does not have as good power as the Box Ljung Pierce test, alright anyway. So, for now this suffices we are happy with the 5th order AR model, but not yet done what are supposed to investigate next with this AR5?

Student: (Refer Time: 08:04)

It is everything is I mean in terms of basic requirement it has understood the; that means, it is not under fit its actually been we do not know now at this stage whether it as over fit or correctly fit right all we know for now is that we have not missed out on anything. It was next test natural one would be to check for over fitting and over fitting manifest typically in terms of large errors in parameter estimates; that means, if the underline process was AR1 even a fifth order AR in that case would have done this for you, but in that case you will see that the estimates and you should do this as homework go back and simulate in AR1 with a fifth order, tenth order and see what happens it will pass residual test, but what you what would you see is large errors in parameter estimates and how do I now come up with the parameter errors in parameter estimates? Unfortunately with this AR 5 model, yes the information is given to you.

(Refer Slide Time: 09:13)

But one has to go through some manual calculations to compute the errors, it does not just give you straight away in a cocked form and that information is stored let us get rid of all mess as if never we made any mistakes.

(Refer Slide Time: 09:34)
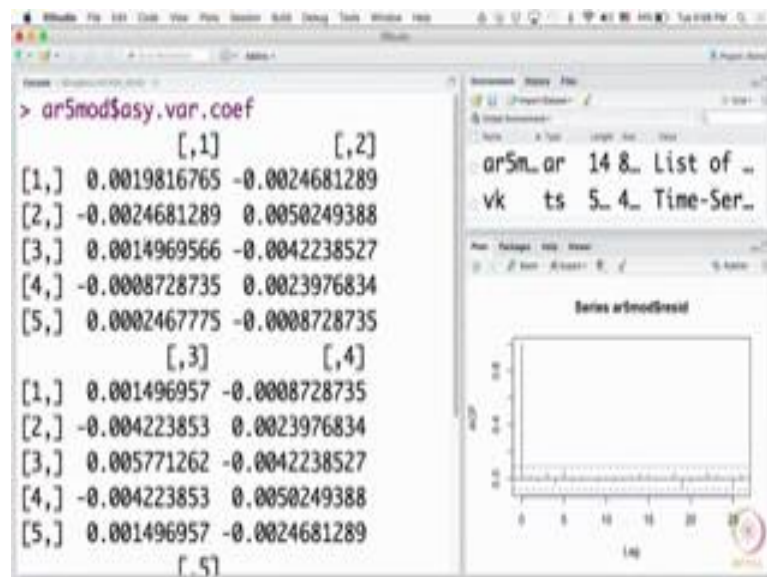
This information on the errors in parameters estimates are stored in the asy standing for asymptotic var dot var standing for variance dot coef coefficients why asymptotic because the expression or the formula there it is using for computing these errors are based on asymptotic theory; that means, large sample theory will again learn that in

during estimation and what it is doing is essentially its says you have estimated 5 parameters because it is a fifth order.

(Refer Slide Time: 10:13)
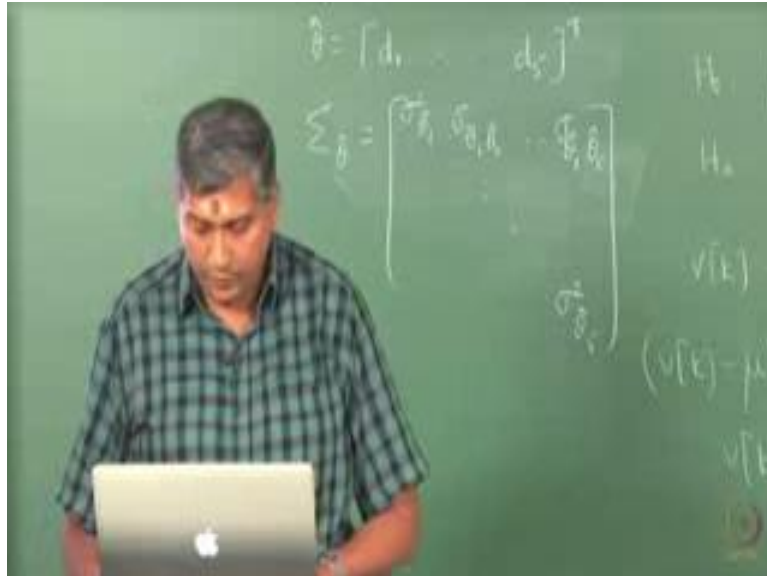
U theta hat consists of d 1 to d 5 and what it is reporting to you sigma theta hat. In fact, its estimating sigma theta hat sigma theta we have discussed earlier in one of the earlier lecture what is sigma theta hat, what does it contain?

Student: (Refer Time: 10:39).

Of the parameter estimates and that is very important because essentially it is a reflection of the errors in your parameter estimate sigma square theta one hat for example, would have would encode or would is a measure of in say. In fact, it is a quite standard error of the estimate theta one hat. So, if I remove the hat essentially here is you are sigma theta what sigma theta hat matrix which contain sigma square theta 1 hat up to sigma square theta 5 hat and then you have sigma theta one hat theta 2 hat and likewise.

Here theta 1 hat theta 5 hat, the off diagonal tell you what is the error that has been incurred in jointly estimating or how the error in one estimate is influencing the error in another estimate all right so. In fact, if you ask for this is gives you a 5 by 5 matrix with the diagonals containing sigma square theta 1 hat sigma square theta 2 hat and so on, typically one looks at only the diagonals it is never going to be a diagonal matrix, but we can just look at the diagonals, we can flux the diagonals from the given one what it is

giving you is an estimate and there are theoretical expression in given by estimation theory on how to compute the sigma theta hat and that theory assumes that is that expression assumes that your model has already pass a residual test, therefore, you should never look at this unless your model as pass the residual test.

Never try to examine the modeling; the parameter estimation errors unless you are convince that the model has at least not gone under fed, as I always say when you are feeding someone, the first thing you want to make sure is that the person is not left hungry once the person has been served enough then automatically of course, guest will tell you whether it is more or enough; in this case its data, so you have to actually make it speak. So, here we should this step should only be invoked an after your convinced that the model has pass the residual test that is very important point should keep in mind.

Now, what we shall do is since we want the errors standard errors and standard errors are the square root of these elements, here we will take the square root of the diagonal, it is too much to write. So, let us actually write here sigma t sigma, this is our sigma t and then we ask for square root of diagonal of sigma t, good. So, these are the errors that you see for this respective parameter estimates, but these errors have to be looked in conjunction with the estimates all right and let us pull up the estimates now. So, these are the estimates that we have and the general result is that the estimate if it is theta I hat theta I hat plus or minus 2 or 3 sigma hat should not contain 0 because theta I hat plus or minus whatever 2 theta hat or 3 theta hat depending on how wider interval you are looking at is the confidence region for the true parameter – meaning, you believe that the truth is somewhere in that interval which is more likely which is less likely we do not know that comes from your hypo estimation theory statistical inference and the hypotheses testing lectures that your probably viewing in that I have already spoken about confidence regions. So, this confidence region is a very very important thing that has to be constructed post estimation and that allows us to test the hypotheses of the form whether theta hat is theta i is truly 0, for all i.
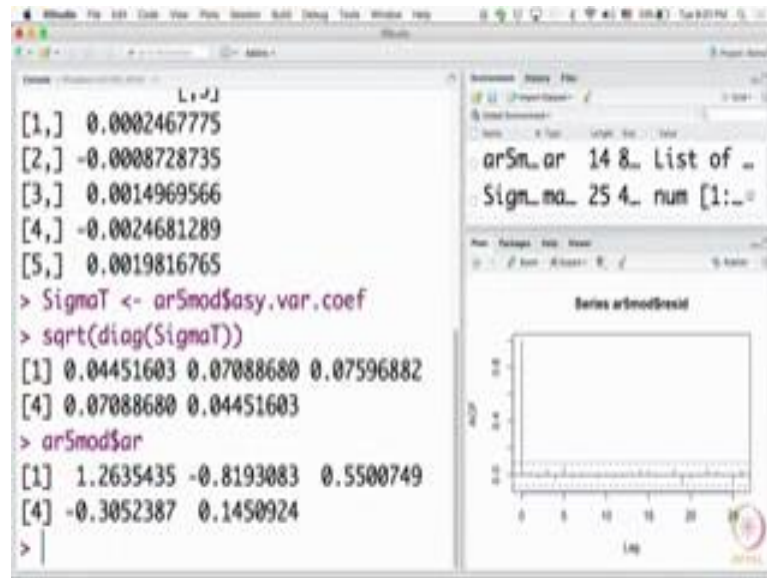
What do you think when you look at the estimates and the errors do you think that any of the confidence regions for the individual parameters contains 0, if it contain 0 then 0 is also possibility and therefore, you fail to reject the null hypotheses that theta is 0 for that parameter. In other words here I look at this value here and then do a let us say I am constructing a 99 percent confidence region or lets a 95 percent it is a bit more accurate if I use to. So, 1.236 that is a 1.264 plus or minus 2 times 0.045, would it contain is 0? No and you can now apply this same procedure to the individual parameters parameter estimates and see if any of the confidence region contains 0, do you notice anything there? We do not, right.

Which means all the parameter here that I have included in the model are significant, I cannot I have not over fed. So, kind of a hit the jackpot all right; that means, the guest that I have gotten from pacf is very good, but sometimes you may say oh no; that means, may have been a guess the residual I can even further whiten it I can you are paint advertisements show each of them complete to show that their paint is more whiter than the other one it almost is unbelievable. In fact, many times I do not know if you noticed the start with the original one and the for example, detergent advertisements they washed one is more whiter than the original one.

I do not know in a bit to actually do excise that is call over fitting, over cleaning anyway. So, you as an exercise you should go back and see if excessive fitting really leads to parameter estimates that are insignificant statistically yes.

Student: Sir if there are reason for this symmetry of the (Refer Time: 17:20).

Yes, that you will understand then we go through the theory there is a reason you will see that symmetry.

Student: Sir, but how does the actually find out the radiance have given in parameter.

Yeah that is where you have to wait until the estimation theory is dot, we go through certain expression and that are some assumption that are made and the starting point is a expression for theta hat; that means, remember we have use for example, Yule Walker's method. So, that is a starting point to you say theta hat is your big covariance matrix inverse times that vector you recall the Yule Walker equation, is an now given this way of expression estimation we know that of course, theta hat is a random variable and then you ask what is a variance of theta hat. So, there are some complicated expression not extremely complicated some complicated steps finally, sigma assumption and large sample theory being invoked as n goes to infinity what happens in. So, and then like formula comes out for your variance of theta hat. So, it bit early now to explain that we need to learn the elements of estimation theory before we go into that.

But as I said it assume the most important assumption that it makes is that the residual survived strictly it assume that you are captured more or less the predictable portion of the model of the series. So, very quickly now let us actually complete this excises and build an ARMA model; now we have not answer this question yet. So, for example, here if you want to know what is alpha it does actually give you as I said in x dot mean. So, this is a value of alpha it is got not may be if they change its now it gives you mu, but never the less unfortunately in the AR routine there is no way of at least it does not report the error in this estimate. So, it is not giving you just alpha or mu whatever it is which giving estimate of those, but what is error in that?