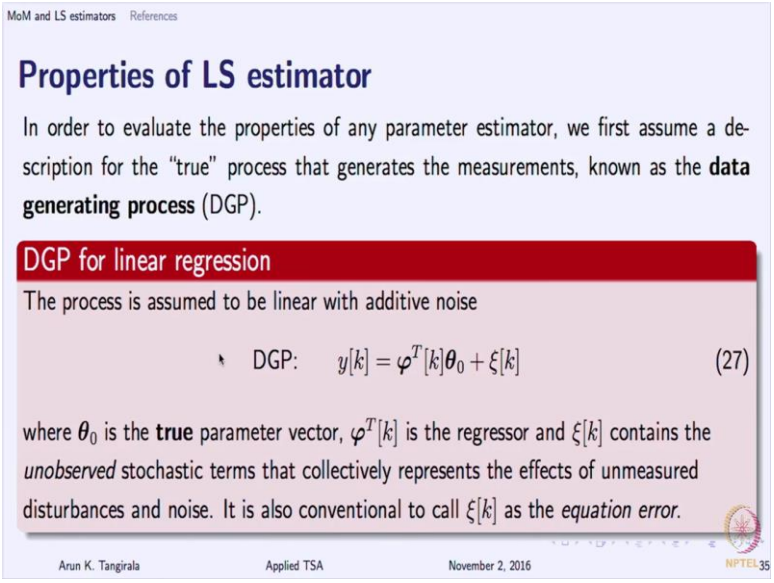
Applied Time-Series Analysis
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture – 100
Lecture 43B - Estimation Methods 1 -7 (with R demonstrations)

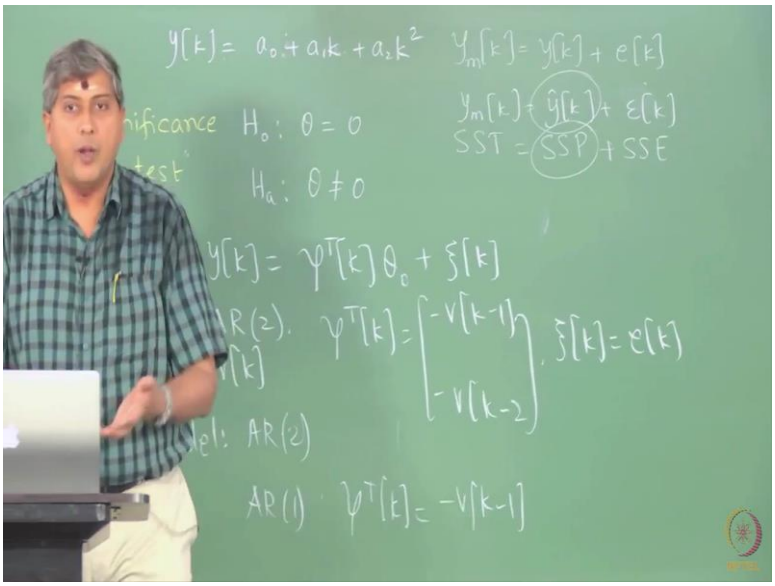
Let us quickly now get back to the theory.

(Refer Slide Time: 00:15)



And as I said yesterday in evaluating the goodness of any estimator we have to describe what is truth. We have to state the truth clearly then only we can talk of bias, consistency and so on. So, we say here that the truth as I have written there on the screen.

(Refer Slide Time: 00:36)

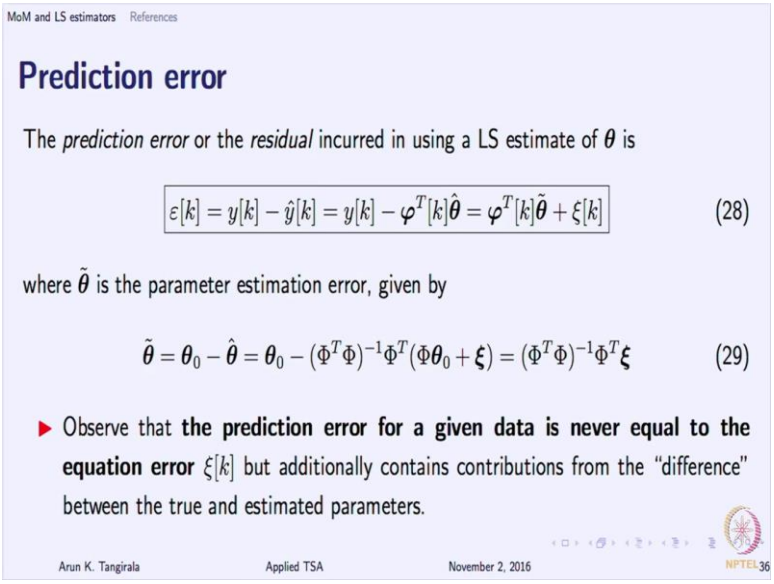


The truth or the data generating process is of this form; yet it is assumed that the process as a same set of regressors in your model. Now what is z? Of course, depends on your regressor set. This regressor set is assumed to be the same as your model, which means that in your model you have captured the right kind of regressors.

If you have not; that means if there is a miss match between the regressors that are truly generating y and the regressors that are sitting in your model then the excluded regressors would go and sit in z. As a simple example we went through yesterday, suppose the DGP is AR 2 then we know that the regressor set consists of your two past outputs and this is your e k. But this is provided; your model is also AR 2 I mean if you are comparing with the model. Suppose I have a model AR 2 then the regressors are the same.

But suppose the model is AR 1 in which case the regressors in the model. So, these are the once that are generating the series y or we call this as v k here for you, whereas when you are looking at the model particularly AR 1 model then the there is only a single regressor. In which case viewed from the model what is left over is e k minus d 2 v k minus 2; we will come to that when we discuss the example I will talk about it.

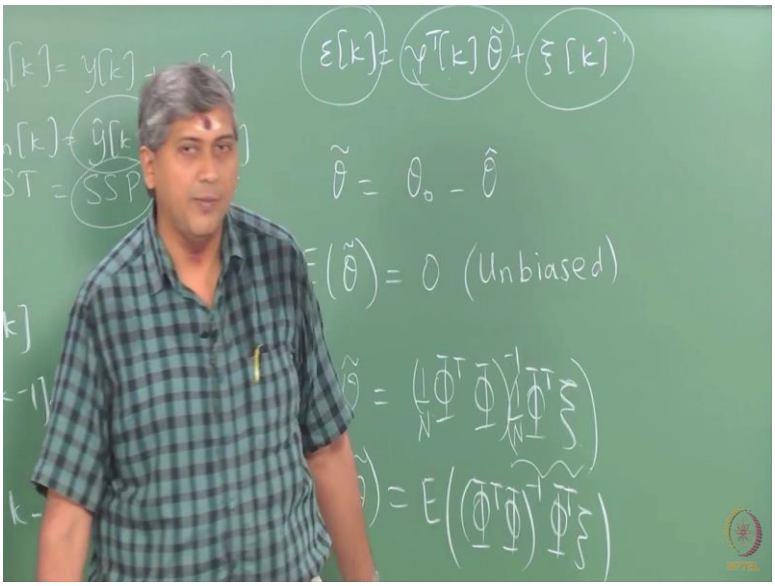
(Refer Slide Time: 02:40)



So, let us for now assume that the regressor set in your model and the regressor set in the data generating processes are identical. In practice, we would never know that the best way to ensure that is to make sure that, in fact you will see that least squares estimates are good so long as these errors which are known as equation errors are white.

As long as you ensure that whatever is left over from your model is white which is done through a residual analysis your safe, in the sense you can expect good estimates from your least squares methods. Now, in order to discuss the properties as I said- the foremost variable of interest is a residual because we have just said residual analysis is a key. So, let us look at the theoretical expression here, the residual is simply y k minus y hat of k this is nothing mysterious about it. We have just rewritten this residual as psi transpose k theta tilde, where theta tilde is the difference between the truth and estimate plus your equation error.

(Refer Slide Time: 04:01)



So, you should observe that this epsilon k in general consists of two terms. In other words even if I have included the right kind of regressors still I cannot expect to recover the z k; that is the true equation errors, I can never ever recover the actual errors that we are generating the data for me. Epsilon, because the epsilon the residual contains in addition to z k this term here which is never going to be 0 for finite n.

Now, having said that a significant part if you gotten your regressors right you should expect theta tilde and if you have large number of observations you should expect theta tilde to be quite small as you will see shortly. In other words if you have large number of observations you would have obtained decent estimates of the parameters, therefore the error in the parameter estimates is going to be low and predominately your residuals will contain the equation errors. And that is good news because we will see shortly that in order to estimate the variance of this part we will use epsilon k.

Why do we need the variance? It become clear shortly; now the object of interest of interest to us is theta tilde, because we know that for example, if I want unbiased estimates you can define it either way theta hat minus theta naught or theta naught minus theta hat we want the expectation of theta tilde to be 0 if I want unbiased estimates. That is a first property of interest if you recall from the properties of estimators. And equation 29 actually gives us the expression for theta tilde from where we derive conditions for unbiasness of least square estimates.

Again, you should ask yourself what is meant by unbiasness; unbiasness is when you have many data records and you are estimating parameters for each of this data record, take the average of all of those estimates you should recover the truth. If you cannot; and there is difference between the average of the estimates across records and the truth then we say it is the biased estimate. So, you can see the expression for theta tilde is phi transpose phi inverse times phi transpose c.

Clearly the properties of theta tilde now depend on three things: one is of course in phi itself that is your regressors and the z, but more importantly if you look at this expression carefully the regressors and the equation errors appear in a particular manner. For example, here I can multiply and divide by 1 over n so that I recognize this part here. What would be this? What kind of interpretation can you attach? What is a size of phi transpose z?

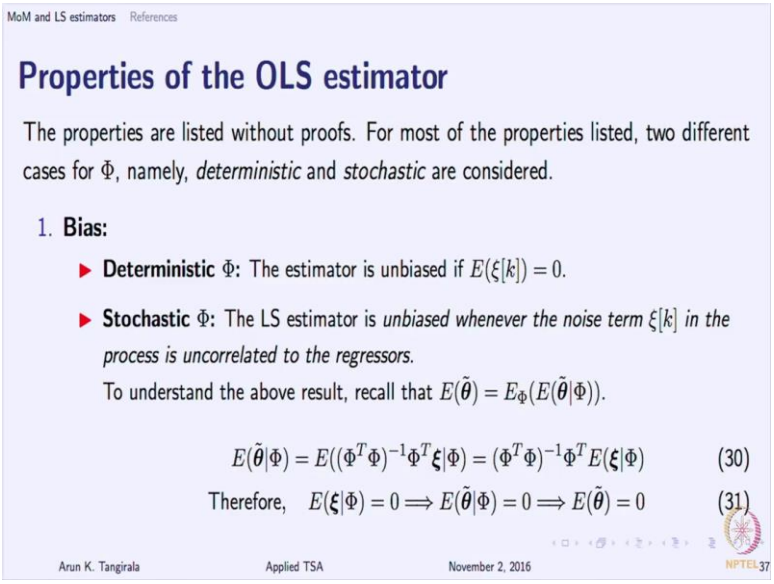
Student: (Refer Time: 07:50).

It is a p by 1. What would it contain the p by 1 vector?

Student: (Refer Time: 07:59).

So, 1 over n phi transpose z assuming that, they are all zero mean and so on would be the estimates of the cross covariance between the regressors and the equation errors. That is whether there is anything common to what has been left out versus what has been included. And we will see shortly that if there is correlation between these two we cannot guarantee consistent estimates or even now we will also end up with the efficiency and bias issues. Let us see that mathematical.

(Refer Slide Time: 08:39)



So, here we have the first result which is on the bias; and always there are two possibilities for the regressors. In the earlier example that is example that I have demonstrated in r what did our phi consist of, just this morning now when we went through an example.

Student: (Refer Time: 09:00).

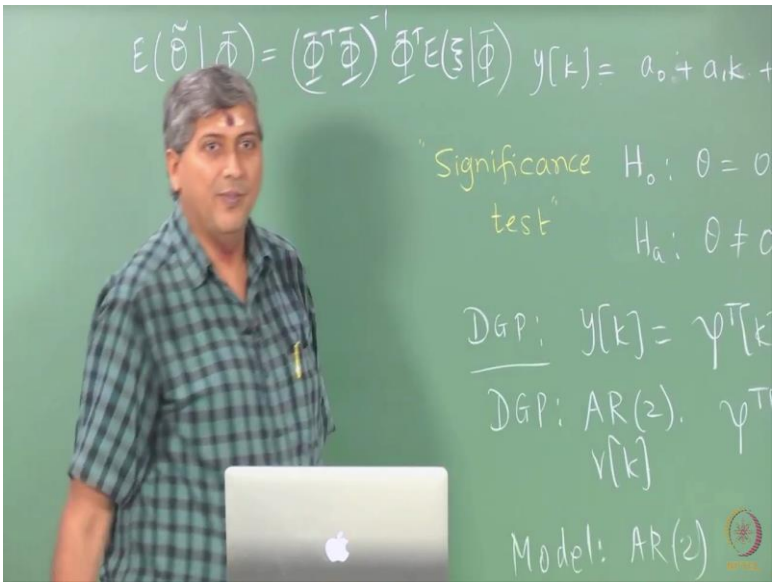
The time stamps, simply the time stamps? There is no uncertainty about them. So, the regressors were deterministic, whereas if I look at the AR model estimation using least squares the regressors are stochastic. In general you can have the regressors being deterministic or stochastic or a mix sometimes, we will not talk about the mixed case typically that arises in system identification, but the point is your regressors can either be deterministic or stochastic.

Now, if you take the deterministic case the condition for obtaining unbiased estimates is very straight forward, because you can look at his expression and this condition. So, if I take the expectation of theta tilde since phi is deterministic all that is required for unbiasness is that the equation errors that are generating this data y they should be zero mean that is all. If they are not zero mean then it says that you will see a bias in the estimates, but that is very mild restriction. All you have to do is make sure you include an intercept term in your model and then you are the things are take care of. That is if

you do not know a priori whether the errors in your data as zero mean or not all you have to do is just include an intercept term.

The case is a bit more complicated for stochastic regressors. In the case of stochastic regressors evaluating the expectation of theta tilde that is expectation of phi transpose phi inverse phi transpose z is not going to be easy, this is not going to be easy. When phi was deterministic I could simply take the expectation passed those regressor matrices, after all they all linear operations. Remember the reason that we were able to take the expectation passed these operations here for the deterministic cases because they are all linear operations and expectation operators are linear operator. But now that phi is stochastic I cannot really take compute the expectation so easily. Normally, what is done is we would compute what is known as a conditional bias.

(Refer Slide Time: 11:33)



So, we ask for a fixed regressor set, when can I expect that is I keep my regressor set fixed and now I am looking at the bias where is the question of bias then if I am fixing the regressors still my z is allowed to vary that is a random signal. So, I keep the regressor set fixed even if it is a stochastic, I will keep the regressor set fixed and let only z vary.

In such cases what is the condition on the data so that I get unbiased estimates. Now it turns out if you just work out the math. Since, we are fixing the regressors we go back to the previous case, but the only difference is now you have phi transpose phi inverse phi

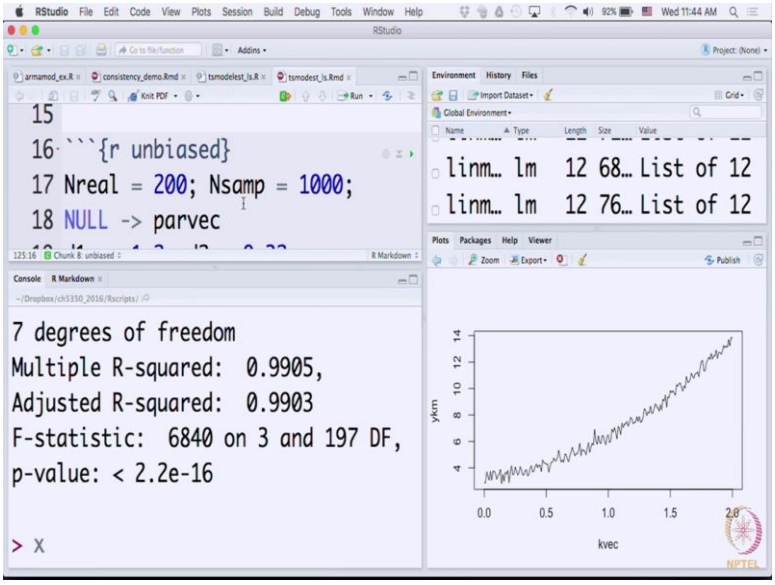
transpose expectation of not unconditional expectation, but conditional expectation. And when this conditional expectation is 0 you can expect unbiased estimates. But what does conditional expectation 0 amounts to? Suppose, I say expectation of x given y 0 what does it tell you about the nature of relationship between x and y? As a independent.

Student: Uncorrelated.

Uncorrelated; so this is what I mentioned earlier. When the errors in your data and the regressors are uncorrelated then you can expect unbiased estimates. Why I am a so worried about unbiased estimates? Because I do not want a systematic error in my parameter estimate, bias always means a systematic error.

So, let me just quickly show that to you in r on the AR model part.

(Refer Slide Time: 13:35)

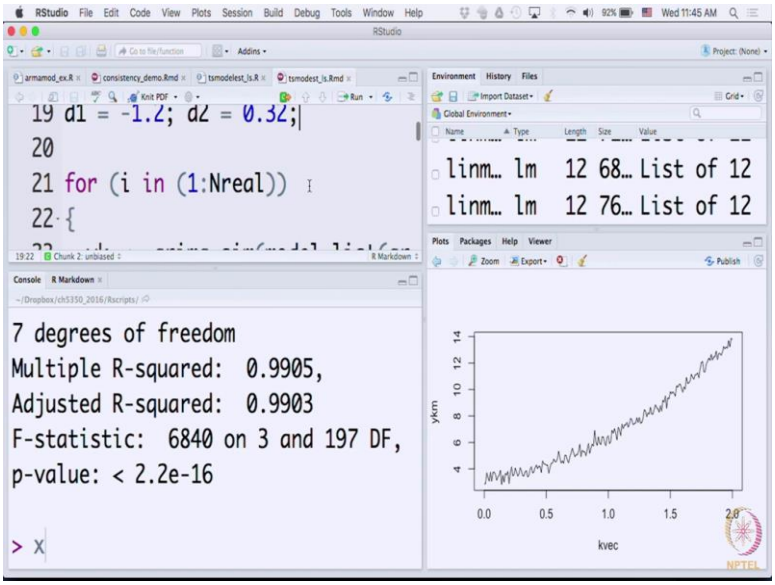


So, I have actually written up here I have marked down document that I will up load on the website.

So, let me go to the top of the document. The first thing that we will demonstrate is when that the least squares method produces unbiased estimates when there is no model miss specification. What we mean by miss specification here is the regressor set in my model and the regressor set in the process are identical. Identical meaning not value wise necessarily, but structurally they are identical.

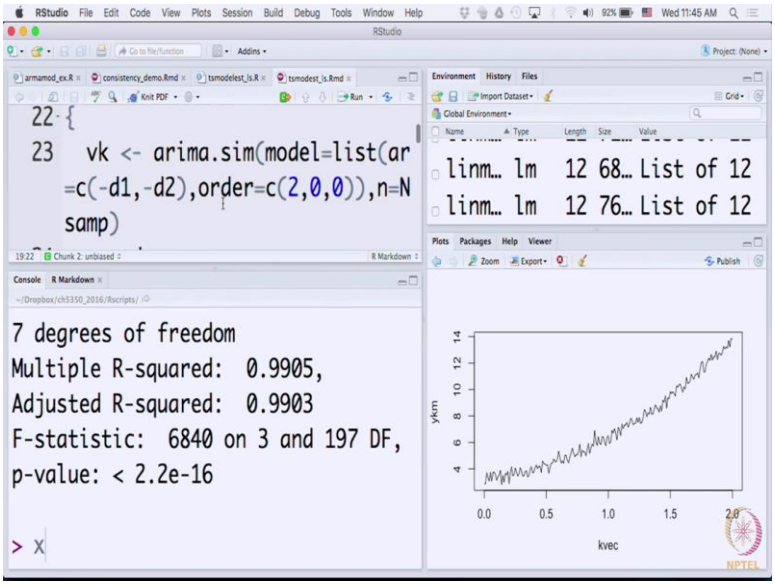
So, to do this yesterday I said if you want to verified bias by simulations you have to generate data repeatedly compute parameter estimates take the average. That is what I am doing here. I am generating 200 realizations, once I load the document you should play around with these values. And in each realization I have 1000 observations. And I am initializing the parameter estimate vector to some null vector.

(Refer Slide Time: 14:40)

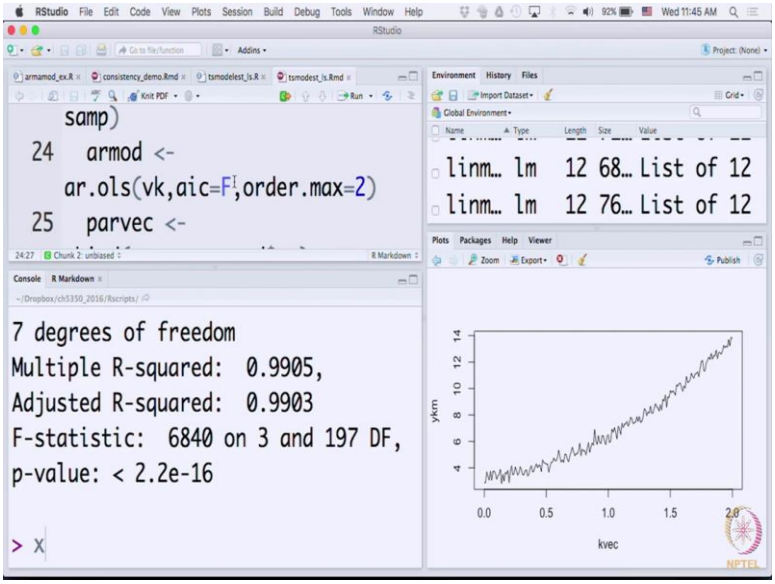


And these are the coefficients of my data generating process. I am going to generate data from an AR 2 process.

(Refer Slide Time: 14:54)



(Refer Slide Time: 15:03)

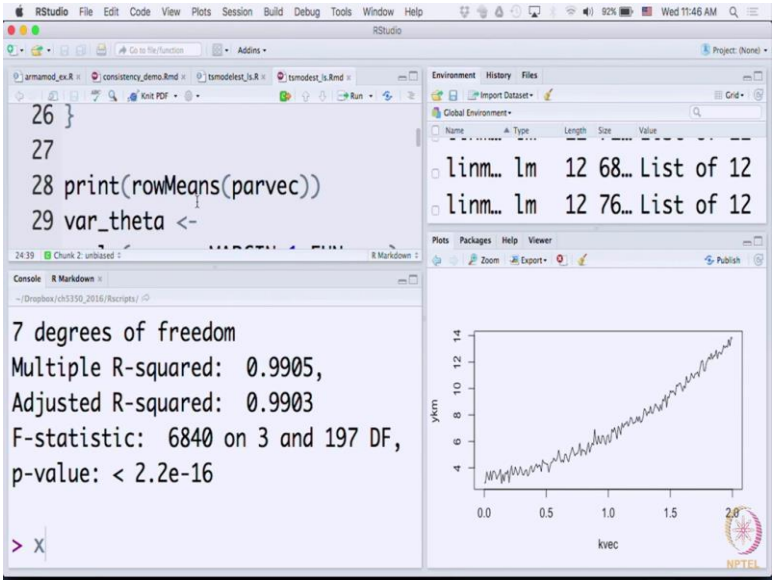


And I am going to repeat this how many times, 200 times. So, here is the data generating process I am supplying the coefficients and the order. And what I am doing is I am fitting now a model of order two, an auto regressive model of order two using the least squares method in r you have AR dot ols you have AR dot burg and then you have AR dot Yule Walker; there are many methods for estimating AR models of which we are interested in ols.

So, now look at a syntax here; I am supplying the series of length n and I am setting this aic option to false when I said this aic option to true then what are this AR dot ols does is it is scans through AR models of different orders up to the maximum order you specify and picks the model with the lowest aic.

But, here I do not want it to do that. I know I want it to actually fit exactly the second order, therefore I said turn of the aic option and set order dot max equals to in which case if it only the second order.

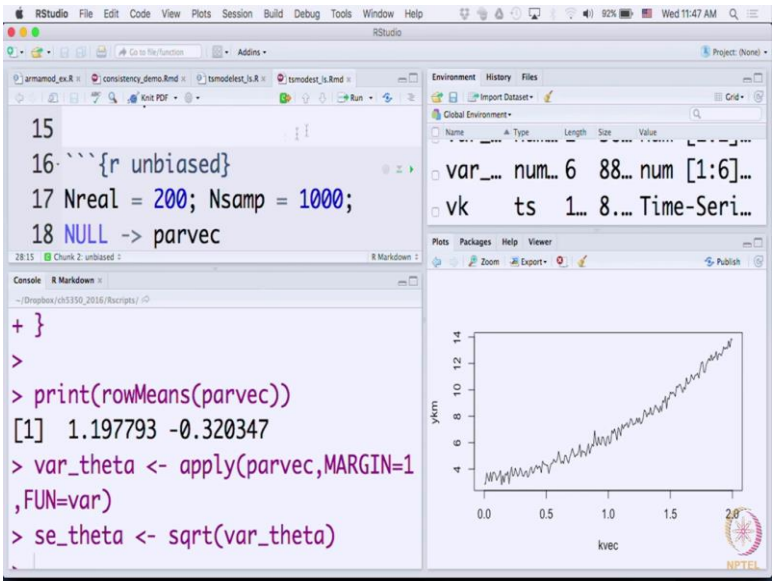
(Refer Slide Time: 16:09)



And once I do that I am concatenating the parameter estimates each time for each realization, so c bind is doing that for me. That is it. And then what I have to do is I have to calculate the average of parameter estimates across realizations. That is what rho means is doing for me.

So, let us actually run this and these chunks of code in the r marked down document and see what it as to tell me.

(Refer Slide Time: 16:36)



So, first let us look at this output here. What is it giving me here? It is giving me the average of d 1 hat and d 2 hat across realizations. What are the true values that we have used? We have used minus 1.2 and 0.32.

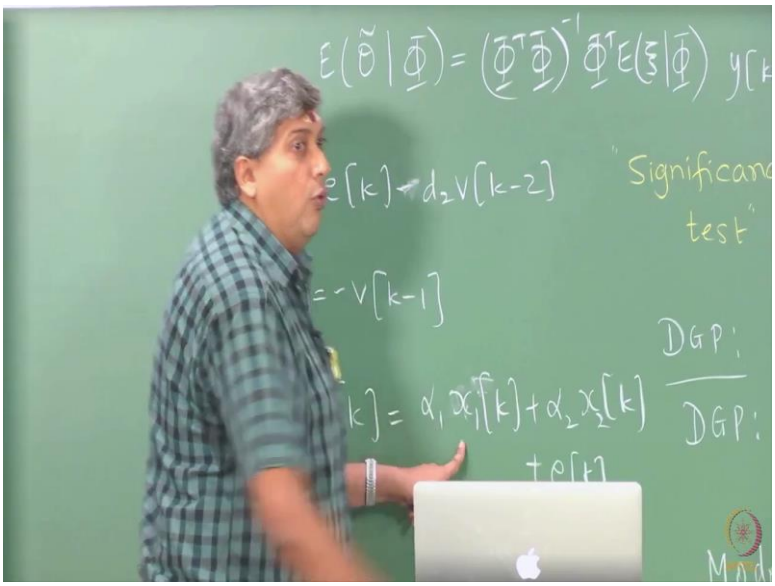
But remember the AR coefficients that AR dot ols returns is a negative of it, because it is looking upon as negative sign coefficients. If you are confused we can in order to compare we can always ask it to print minus of this so that when I do this there is no confusion, it is just the difference in the syntaxes that we have been a notation that we have been using. Here you see minus 1.2 nearly and 0.32. If you increase the number of realizations will be more and more accurate. In this case therefore, can I confirm that I have obtained unbiased estimates this is how you should confirm, and this is the typical way of testing for unbiasness through simulations.

Now, let us quickly move on to the case. There are other calculations that I making, but I would not come to that now. Let us now actually generate, the distribution also will skip for now; we will check for first bias and then come back to consistency. So, what happens is suppose I fit an AR 1 model alone, when I fit an AR 1 model what is my regressor this is a regressor, what have been left out?

Student: e k.

E k plus d 2 v k minus 2; and what theory tells us is I will get unbiased estimates if and only, well we have not proved the if and only if; if whatever I have left out is uncorrelated with whatever I have included.

(Refer Slide Time: 18:54)

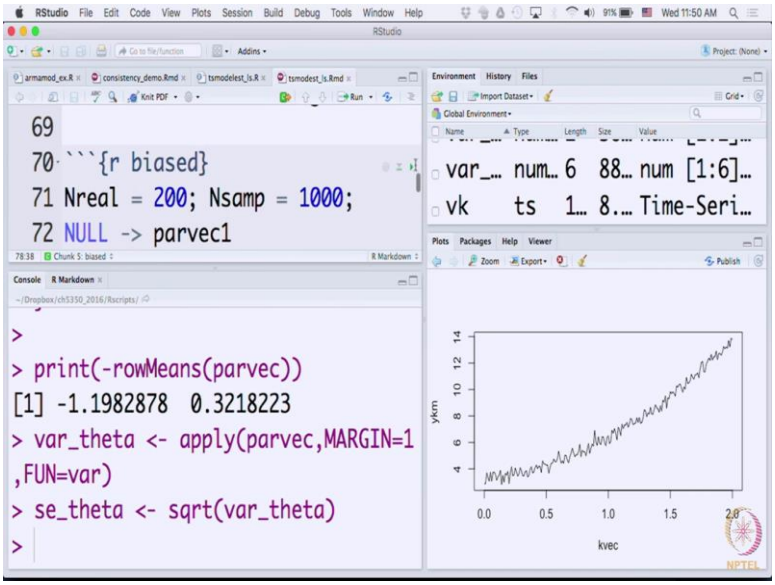


And whatever I have left out is e k plus d 1 sorry d 2 minus d 2 v k minus 2 this is what I have left out. And what I have included is v k minus 1 or minus v k minus 2. So, this is we can say whatever I have left out and this is what I have included in my regressor set. Are these two correlated? What do you think?

Student: (Refer Time: 19:24).

Therefore, we should expect a bias estimate of the coefficient d 1. And the question is does that happen.

(Refer Slide Time: 19:36)



So let us just run this code very quickly here; the same story except that now I am fitting an AR 1 model, otherwise if the rest of settings remain the same. So, if we were to run this chunk of code here this is the averaged across realizations. The true value is of course once again here there is sign is reversed, but the true value is 1.2, whereas I obtained 0.9 that is a big difference, you cannot say the difference between 0.9, 1.0 is only due to a chance, [FL] there is nothing like that. These biased and theory tells you it is by unbiased.

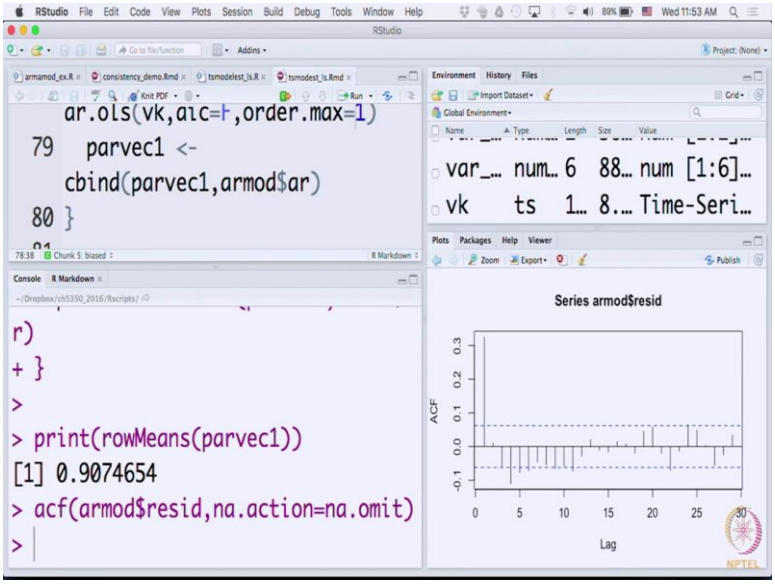
As a simple exercise what you should do and you should note this down; go and generate another process not necessarily an AR process but some synthetic process where you generate data of this way y k equal some alpha 1 x 1 square sorry x 1 k plus alpha 2 x 2 k plus some e k. Where, x 1 and x 2 are some random signals they are uncorrelated. Step one you generate data of this form, x 1 and x 2 are some random signals uncorrelated random signals. And then you fit a model, so this would be your data generating process; then you should fit a model including only having only one regressor throughout the x 2. That means, do not include x 2 in your regressor set a s a result what would be left out is e k plus alpha 2 x 2 k. See if you get unbiased estimates of alpha 1, what do you expect?

Student: (Refer Time: 21:26).

Yes, because x 2 and x 1 are uncorrelated. But that did not happen in our time series modelling because, these x 2s and x 1s are related and that is why you have a problem. In this case even though you omit this regressor there is no issue you will still get unbiased estimates of alpha 1. So, that is what is the meaning of that unbiased result there. You will incur a systematic error in your estimate if whatever you have left out is correlated with what you have included.

How do you fix this problem in AR 1? Look at residual analysis and first you can look at residual analysis to what do we expect we expect white residuals or what you can do is to check if I to explain why I get unbiased estimates I can also compute cross correlation between residuals and the regressors what I have included. But typically in time series modelling simply looking at the residuals alone is enough, what we mean by residuals is ACF of residuals. We will tell you that the residuals are not white. And therefore you should go back and correct your model.

(Refer Slide Time: 22:45)



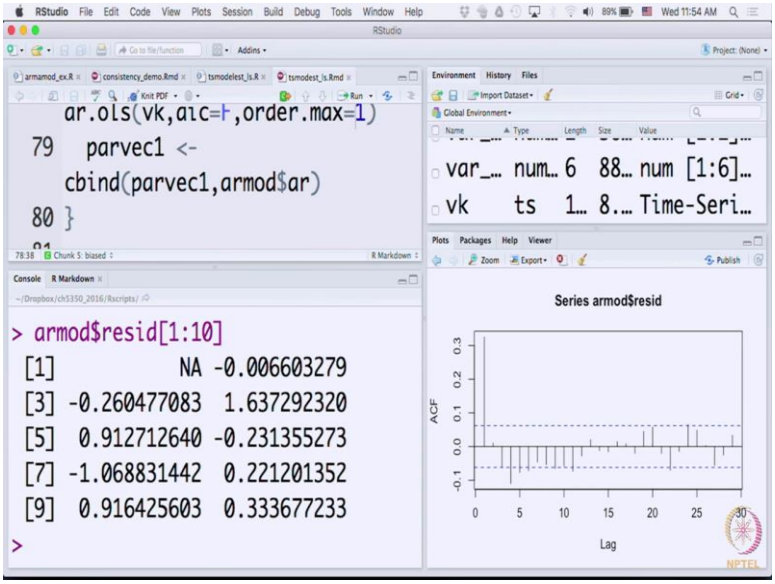
So, if I look at the ACF of; here we are fitting let us say pick any AR model here armod dot resid and you have to make sure here that the NA values.

So, you can see here that the ACF of residuals shows some significant correlation. The reason for supplying this NA dot action is I have explain to you before, when you compute yesterday we talked about is when you set up the regressors for AR modelling you start from.

Student: 2.

2, if it is AR 2 if it is one you start from 1, but the residuals are calculated from 0 onwards.

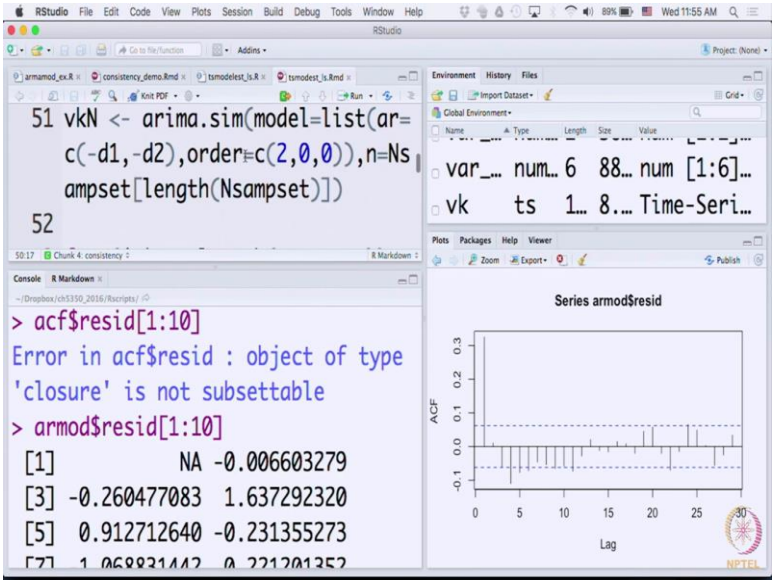
(Refer Slide Time: 23:32)



If you look at the first value of the residuals here it will show you NA at the first one because there is no prediction there. What we are telling is ACF is to ignore that NA value that is that is something that you have to keep in mind. ARIMA on the other hand will not result in and these things. The final thing I will just demonstrate is consistency, we will go back to the correct model and then we will continue our discussion tomorrow.

It is important to understand the simulations as important it is to understand the theoretical result so you should pay attention to this. So, let us go back to the case where I correctly specify the model order, then only I can talk of consistency and so on. If there is a miss specification then there is a problem. So, here what I am showing in this chunk of code is consistency. All I am doing is I am actually generating only a single realization now, remember consistency looks at how the theta hats behave as you change the sample size. So what I am doing is, I am generating data of a big size but I am only taking sub sets of it. I take for example sub set of sample size 10 then the next one, so I increment here.

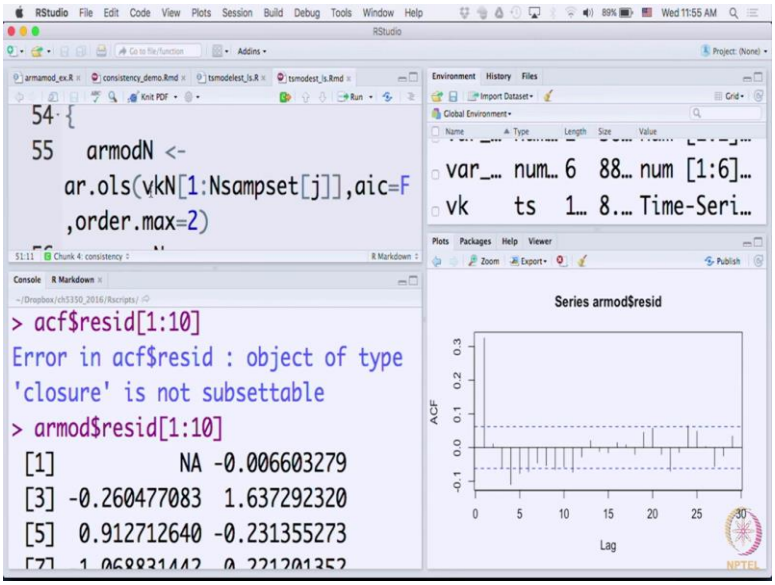
(Refer Slide Time: 24:59)



So, you can see here the sample set is 10 power 1 and then 10 power 1.25, 10 power 1.5 up to 10 power 6.

And the model remains the same. So, here is the theta hat n this parvec n is theta hat n and the process is being simulated.

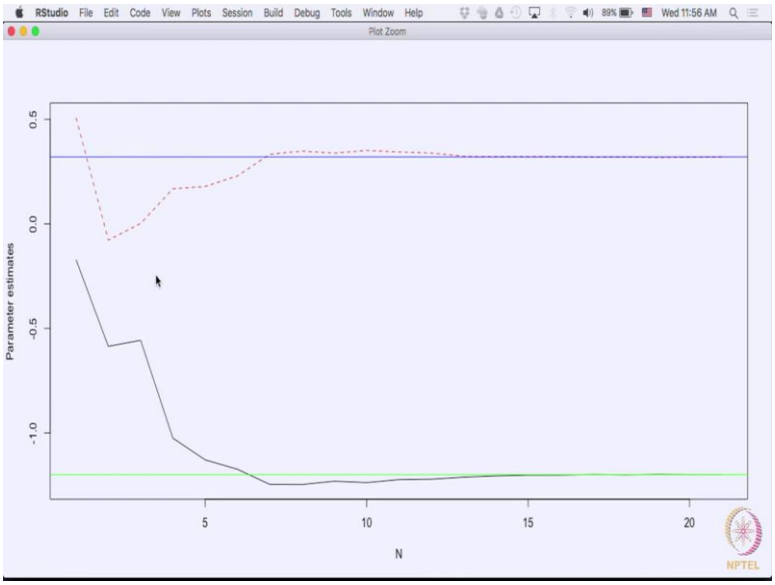
(Refer Slide Time: 25:20)



And all I am doing is I am calculating theta hat for different sample sizes. When I run this code and plot the resulting estimates as a function of m if it is consistent the estimate

should go and converge to their true values as I increase n. So, you can see here what is happening.

(Refer Slide Time: 25:38)



So, there are two plots here corresponding to two parameter estimates, and the blue and green lines are the true values of the parameters. One is 1.2 and minus 1.2 and other is 0.32. So, you can see that the respective parameter estimates I do not how well you can see on the screen, but this is the 0.32. And on the x axis we have sample size, as I increase the sample size the parameter estimates actually, go and sit at the true values whether it is point the d 2 or d 1.

When this happens for any estimator then you can be assured that they are giving you consistent estimates; this is what is consistency. Again, I am showing you this because sometimes it may be very difficult to understand the theory, but in simulation it is a lot more clear and obvious us to what is happening. Tomorrow what we will do is we will quickly wind up least squares, weighted least squares and not linear least squares, where I will talk of distributions and also give you theoretical expressions for what I have calculated here and then that will be the closure of least squares.