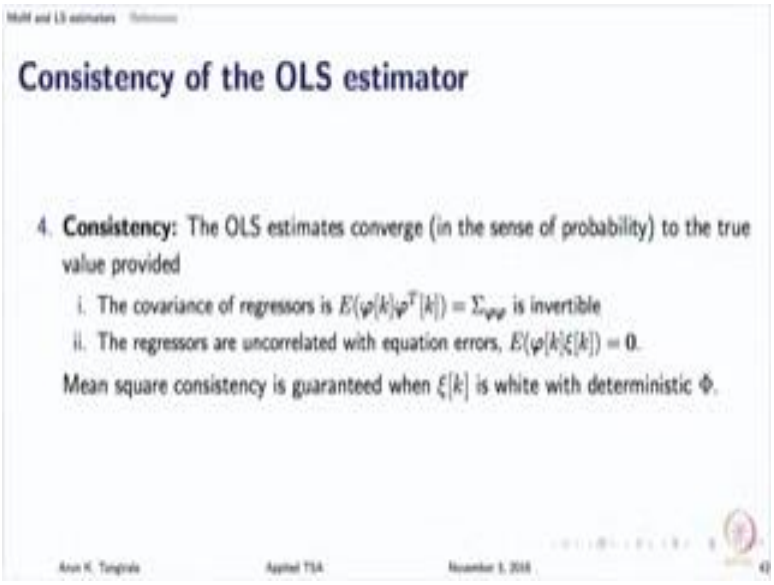
Applied Time-Series Analysis
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture - 102
Lecture 44B - Estimation Methods 1 -9

The next of course is consistency remember we said bias variance efficiency consistency and finally confidence regions. Now we avoid the proves here and I give you the results straight away.

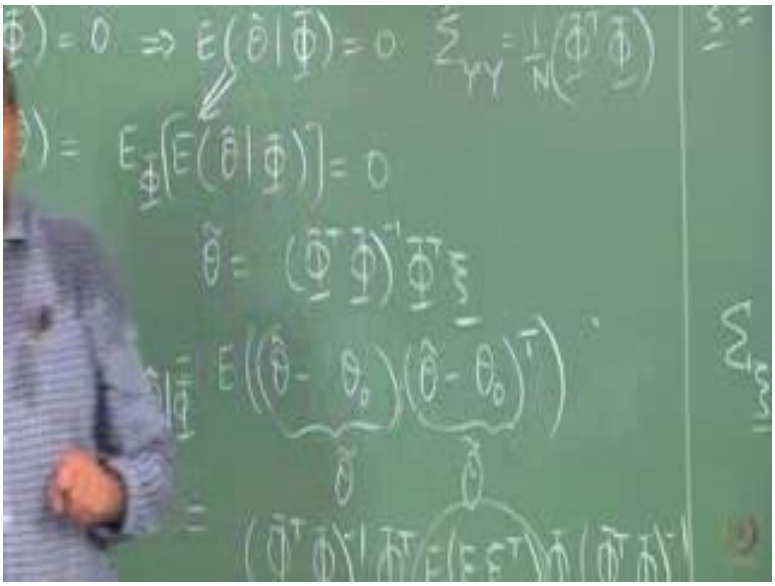
(Refer Slide Time: 00:24)



The consistency of OLS is achieved under two conditions: one is that the regressors, the covariance matrix of the regressors is invertible.

Now, I am giving you the result on the theoretical covariance that is expectation of z k times z k transpose, what this means in practice is this phi transpose phi that you here that you see here it should be nonsingular.

(Refer Slide Time: 00:56)



Because remember and estimate of the theoretical covariance is as I said 1 over N phi transpose phi and just one more point that I want to mention when going back here, when you have n minus p here in the expression 34, all the n's that we are using here refers to the number of observations that is all right, but in practice the numb that n there refers to the number of effective observations that I have gone into constructing your regressor matrix. What I mean by this is when we were constructing the regressor matrix for AR models; we noticed that we had to throw away p observations when I am estimating parameters of an ARP model. So, in effect I am working only with n minus p observations already, right?

I have, I am throwing away n minus p observations, I am working with n minus p. On top of that you have p degrees of freedom lost. The regressor matrix itself is constructed from n minus from n minus p that is it has only n minus p rows not n rows, on top of it I lose p degrees of freedom. So, when you verify the results that I ask you to do, you have to be careful the n here is the effective number of observations that actually go into the construction of phi.

In other words for AR models if you have generated thousand observations and if you are fitting let us say a 3rd order AR model, the n is 997 and p is 3. So, in the denominator you would have 994 whereas, for static models, what we mean by static models is the regressors are instantly related the predictors are instantly related to the regressors, those

are called steady state models, those are the models that you learn in typical statistics courses; where there is no notion of dynamism, in such cases n is the number of observations that you have because you are you do not have to throw away any observations. So, please observe this distinction. So, let us get back here for consistency two things have to be guaranteed, the regressors cannot be linearly related the regressors themselves cannot be linearly related, why is this condition coming up?

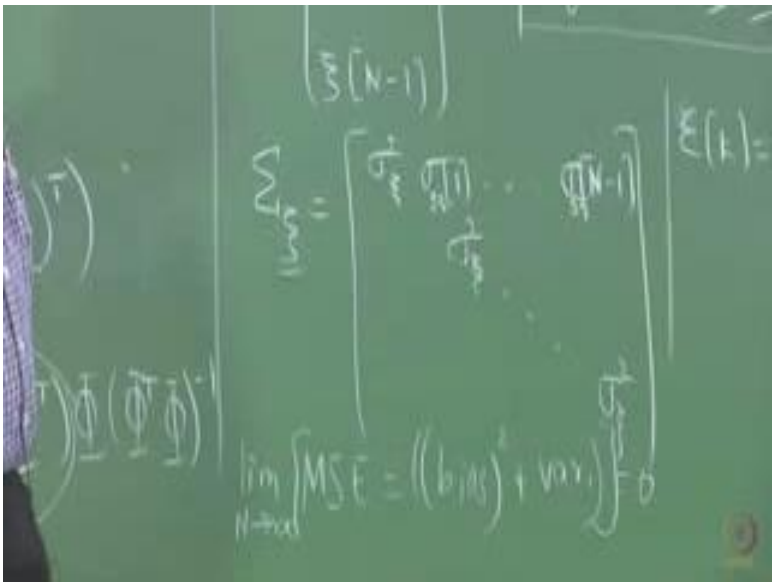
Student: (Refer Time: 03:33).

So, because if regressors are linearly related. So, imagine columns of phi are linearly related, what happens to the rank of phi transpose phi? It becomes rank deficient and therefore, it becomes you know singular and then inevitability cannot be guaranteed. So, collinearity among phi should be avoided and that is a big branch of literature in parameter estimation, particularly concerning with regression models where people are trying to address collinearity in phi, you may not in practice run into perfectly linearly related regressors; that means, any pair of columns of phi may not be perfectly linearly related, but even if they are close then the condition number shoots up and then you have an issue.

So, collinearity does not mean necessarily theoretically it means perfect linear relation, practically it means nearly linearly related. In such cases there are remedies which we do not discuss its outside the scope of this course, but there are remedies available, but what this result tells you is when you are using OLS make sure that your regressor matrix is far away from such situations. Secondly, the most important thing which we have already discussed which is that the regressors are uncorrelated with the residuals.

Consistency looks at asymptotic bias whereas, the earlier bias that we discussed is statistical bias for a fixed (Refer Time: 05:11). So, what this result tells me is even asymptotically the bias will not vanish if the regressors are uncorrelated correlated with the residuals. The earlier case that we talked about the bias is the case when n is finite. So, we are just looking at across the realizations of the equation error whereas, consistency is looking at asymptotic bias, it says the mean square error should go to zero for example, if you are looking at mean square error consistency and mean square error should go to zero in the limit as n goes to infinity.

(Refer Slide Time: 05:48)



And if you recall the mean square error has two components to it. It has bias square plus variance and consistency one form of consistency that we talked about which is mean square consistency, this should go to 0 right. So, which means what the statement implies is that, if the regressors are correlated with the residuals with whatever you have left out in your model. Then this is the term that does not go to zero whereas, variance would go to zero remember that we said in practice we replace variance expression with sigma square E hat and where that would have an n factor which will take care of things, which will take care of the growing phi transpose phi.

Remember as n grows the phi transpose phi also grows because the number of terms that you are summing up is growing. However, there is another n in your sigma square E hat in the denominator, which will take care of that that is one way of looking at. So, the issue is not with the variance per say the issue is with the bias, when the regressors are correlated with the residuals.

So, to summarize when these two conditions are satisfied in the limit as n goes to infinity, consistent estimates are guaranteed. So, go back to the example when I upload the mark down file, download it play around, just make sure in your simulation one of this conditions is violated particularly the regressors being uncorrelated, do not try to play around with the first one because that will bring in collinearity and messes up things

a lot more than you can handle. So, just simulate a situation whether regressors are correlated with the residuals and then you will see that consistency is also not guaranty.

In other words if you are working with ordinary least squares method, just make sure that your residuals are white that is it. By whatever you do you have to make sure residuals are white; that means, either you have to (Refer Time: 08:07) your finally, what you have in your hands is the model. So, trick your model such that the residuals are white, but without running into the problem of over parameterization. So, in parameter estimation there are this walls that you will hit against you have to actually work well within the walls of the framework, that is what is important and the boundaries are different with different estimation methods.

So, finally, we come to the distribution which will tell us which will allow us to construct confidence region. One of the things that you should notice in the expression for theta hat, so let me write the expression here for you again.

(Refer Slide Time: 08:52)

The OLS result; theta hat is phi transpose phi inverse phi transpose y. Assume for a moment that phi is deterministic, just assume that phi is frozen then straight away it tells you that theta hat is a linear estimator that the least squares. So, let me write thus write this for you emphasis that, that the least squares estimator is a linear estimate; why is it a linear estimator? Because it is just a phi is fixed then all your doing is its actually you can rewrite this as sum alpha k, Y k or you can say w k, Y k. So, all you are doing is you

are just linearly fusing the observations that you have to produce the parameter. I will just go back to alpha because w is reserved for weights later on in w (Refer Time: 09:55) I do not want to confuse you; these alphas are calculated from your phi transpose phi inverse phi transpose.

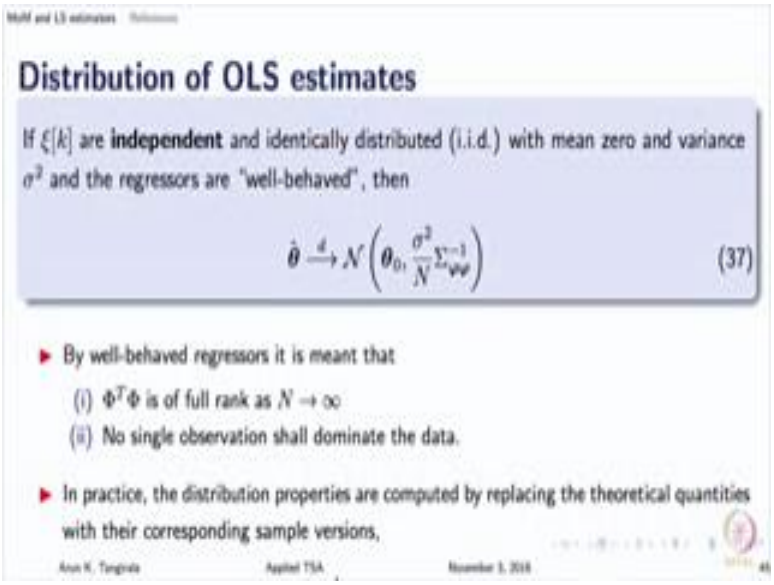
Now, in order to dist determine the distribution of theta hat, all I have to do is invoke c l T right? if y k as Gaussian errors no worries even for finite n theta hat will have a Gaussian distribution, joint Gaussian distribution remember your theta hat is a p by 1 vector, you should remember that. So, although I write it this way you have to be careful let me probably even write this as simply some A times y. So, that A is a matrix that is phi transpose phi inverse times phi transpose that is nothing, but your pseudo inverse. So, you can see straight away that if y has a Gaussian error that means, if you would z is Gaussian distributed jointly, then theta hat is also jointly Gaussian distributed. On the other hand if the data has non Gaussian errors, then only asymptotically theta hat has a Gaussian distribution that is by virtue of the central limit theorem, but that is a fortune that we have with least squares, although there are some restrictions it says it is efficient only when things are white. It is consistent only when the regressors are uncorrelated with the residuals and all of that nevertheless it gives you some benefits somewhere; like in everything else in life I mean if you look at a car, car will have some feature and some features may be absent, but then it depends on what you want.

So, likewise here if you take an estimator it has some features and not you know, but with some restrictions; if you are willing to relax this linear estimation expression, then you will get better properties which is what MLE does. Although MLE is not base on that philosophy, in general all maximum likelihood estimators are non-linear estimators in general may be bearing some special cases; whereas, all OLS solution estimators any for any problem OLS gives you a when the regressor is linear on the other hand with MLE even if the regressor is linear, you would you are not going to get a linear estimator that is the difference.

That is the advantage of working with least squares methods and that is why it is very popular. Imagine now for above 250 years least squares is popular, we keep talking about some DDLJ running in some Maratha theatre for you know 30 years or 20 years and so on, but here least squares has been running in the theatres of parameter estimation for centuries; why? Because of its beauty in a many other things and its simplicity, so

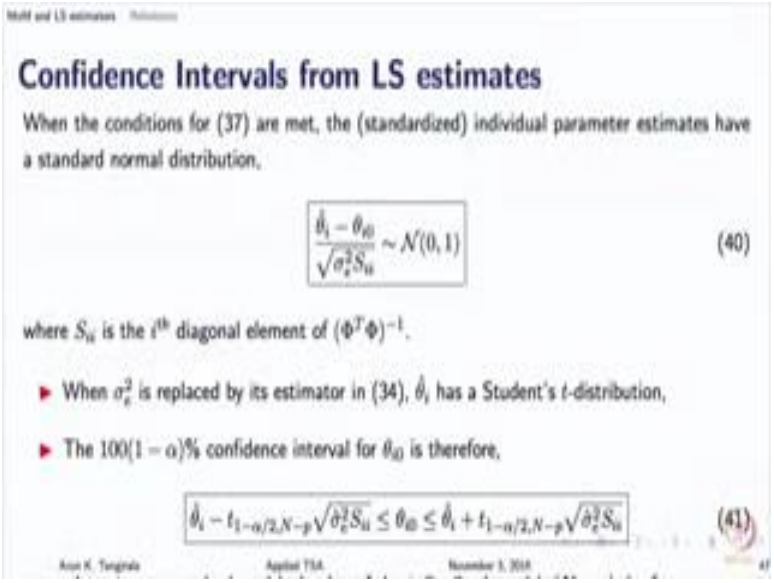
that is it so in general we assume now theta hat to be a following a Gaussian distribution, but keeping in mind that it is a result that holds in general for large n for non Gaussian errors. So, that is the statement here, remember we do not write the statement exactly this way we say rather root n times theta hat minus theta naught follows an asymptotically and follows a Gaussian distribution right.

(Refer Slide Time: 13:16)



Again you do assume that phi transpose phi is a full rank number 1 and there are no out layers in the data or no observations in the data that will hijack the analysis. Few observations after all we know there is no robustness incorporated into the estimator, a there is no robustness explicitly anywhere incorporated into the estimation for into the parameter estimation formulation, as a result a few extreme data points can completely hijack the estimate and that is what we mean by well behaved regressors. Well behaved regressors would mean that no single observation will dominate the regressor matrix and that phi transpose phi is of full rank particularly as n goes to infinity.

(Refer Slide Time: 14:11)



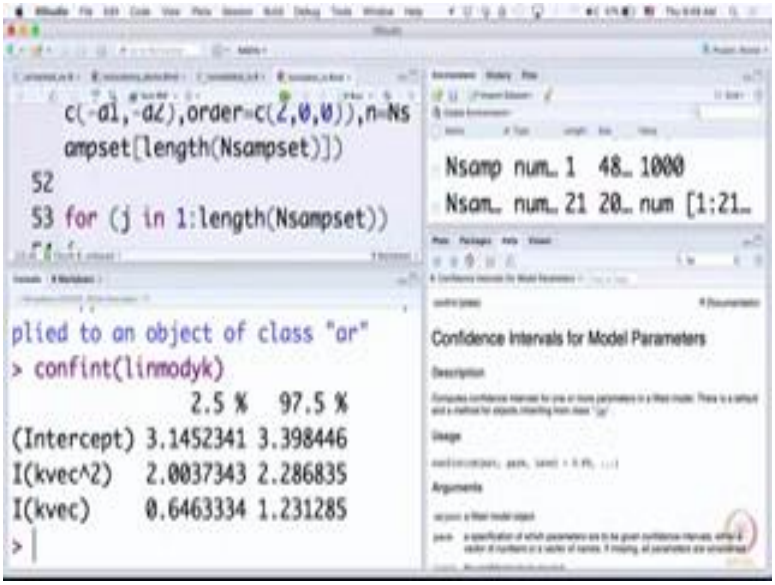
So from this we I will I will skip that part from this we actually compute the confidence regions, these are approximate confidence regions because I m giving this confidence region for individual parameters, the distribution that is given here is the joint one, right? If we know from random variable theory that if I have vector of random variables have a joint Gaussian distribution, it does not necessarily it does mean of course, that individual ones are Gaussian distributed, but I still have to analyze them jointly.

On the other hand if the individual ones are jointly are Gaussian distributed then the joint ones need not be Gaussian distributed, what we are doing here is we are saying is, given that theta hat has a joint Gaussian distribution, the individual ones that is a marginal ones are also Gaussian, but I am ignoring the other aspects that is the correlation between the individual parameter estimates. I am just taking them out and analyzing them individually and saying that theta I hat minus the true value divided by the root sigma square e s i i; what is s i i ? It is a diagonal element of phi transpose phi inverse. So, if you this expression here in 37 is a theoretical one.

In practice what do we do? We replace the sigma c c with this estimate; that is what I do the 37 gives you the theoretical result, theoretical meaning in terms of the theoretical covariance matrix of regressors, in practice I replace the sigma psi, psi with its respective estimate and once I do that you can see that the n will vanish from the expression, so that you have again sigma theta hat being, sigma square e times phi transpose phi inverse. So all that this result is saying is, calculate your sigma theta hat, take the diagonal elements right and then write an individual confidence region.

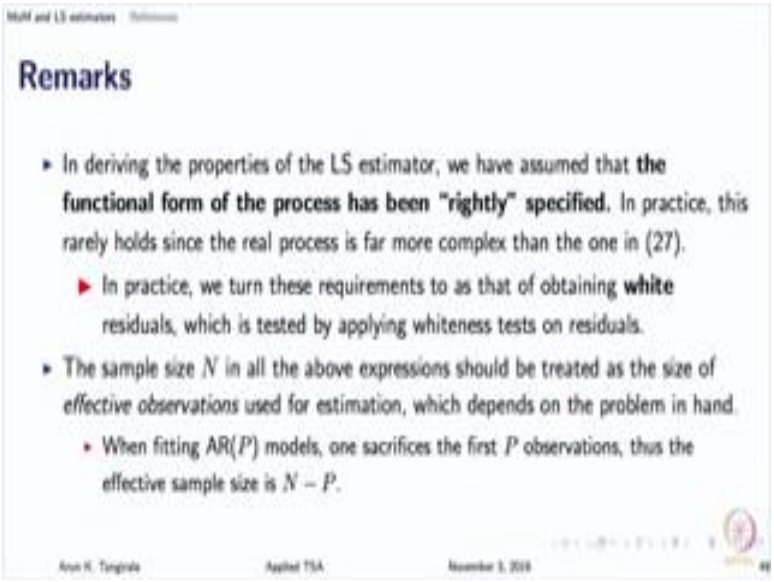
Since we are estimating sigma square e and we do not know sigma square e; the confidence region strictly speaking should be written in terms of T distribution, but asymptotically it is a Gaussian distribution; for finite small n we replace this Gaussian distribution in equation 49, 40 with the T distribution, but that is only valid when the errors are Gaussian. So, to avoid the confusion simply replace the T values there with Gaussian a corresponding Gaussian value distribution values when n is large; typically if in our cases n is large the small sample cases are different.

(Refer Slide Time: 17:21)



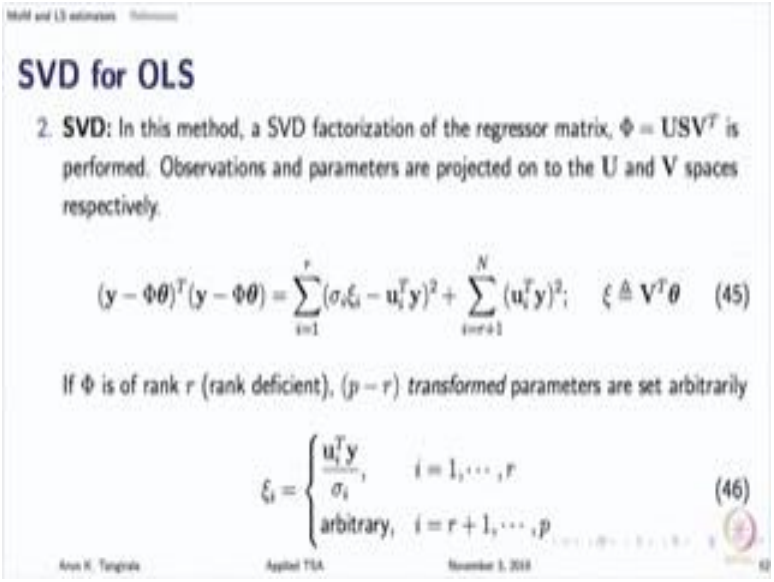
So, that that is how you construct the confidence regions and if you were to look at there is command in r if I am right? Which computes the confidence intervals, so you can supply I do not if it will take this ops sorry; yesterday we had fit and unfortunately this confint only works with l m. So, I if you remember we had computed for example, a linear static linear regression model for y where we did a polynomial fit, you can supply that here and it reports the confidence intervals 95 percent confidence interval, you can change that. This confint does not necessarily work with all kinds of objects, only the results coming from l m can be effect to confint, but there are other packages which will do a better job for you. So, it is giving you the confidence interval, which you could have calculated as well.

(Refer Slide Time: 18:16)



So, that concludes discussion on OLS and I am going to.

(Refer Slide Time: 18:19)



Skip the computation; computation of OLS I have already talked about it, I have said that either you use q r or SVD.