# Applied Time–Series Analysis

**Prof. Arun K. Tangirala**
**Department of Chemical Engineering**
**Indian Institute of Technology, Madras**

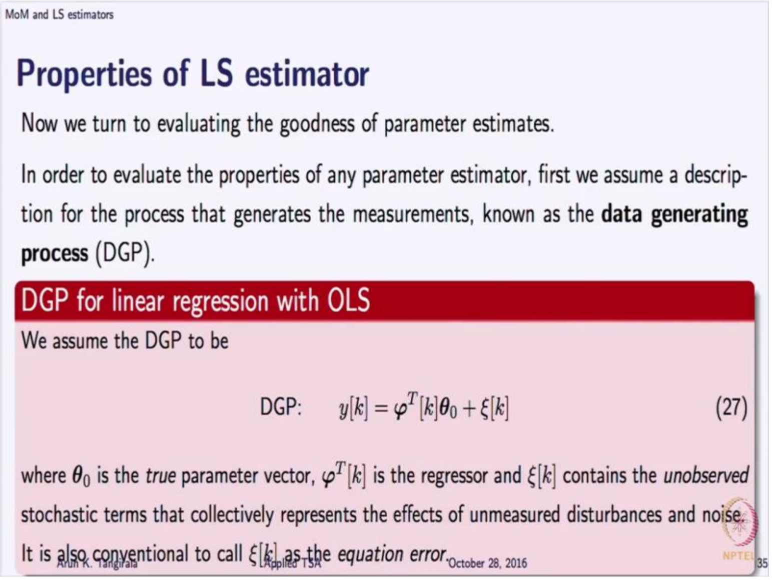
## Lecture – 98
## Lecture 42C - Estimation Methods 1 -5

I will just briefly talk about now the most important question then we will continue tomorrow. So, until now we have talked about how good a fit least squares has achieved, usually by measured by r square and adjusted r square, but these measures do not tell me anything about theta, they do not tell me how good an estimate of theta I have obtained, what do we mean by good? Now we have come to practicality every measurement has some randomness in it. In the wake of this randomness in y have I obtained an unbiased estimate of theta, have I obtained a consistent estimate of theta, efficient estimate of theta these are the three concerns that I have and what is the most important thing that we do, in any estimation exercise what is the final thing that we report confidence regions.

So, these are the four things that we want to ask do I obtain an unbiased estimate of theta, do I obtain efficient estimate of theta consistent estimate of theta and how what is the distribution of theta hat so that I construct the confidence region and to be able to do. So, I have to assume that y is generated as per some law, remember we talked about notion of truth; now I have to fix the truth, I have to say if the truth is this how good is least squares in being able to recover that truth. In reality please remember that the truth is never known, there is no equation that exist perhaps that we know.

But what we are demanding is if the truth is evolving as per some equation at least in those situations does least squares give me good estimates consistent estimates efficient estimates. So, for this reason always in any estimation exercise or analysis assessment of any estimation method least squares MLE whatever we first begin with what is a known as DGP data generating process, we freeze that first you say if the data generating process is this; then now tell me how good your method is.

(Refer Slide Time: 02:44)




So, here we assume that the data generating method a process has a linear structure to it, that is a big assumption, but it is ok; I am solving I am fitting a linear regression I am fitting linear regression model therefore, it is fair to the model only if I assume a linear process, I cannot assume non-linear process and say are you get me the truth; there is no comparison structurally. When we talk about truth there has to be some similarity between the model and the process some similarity and some difference the similarity here is in the nature and the structure. So, we are saying let y evolve as psi transpose k times theta naught plus z.

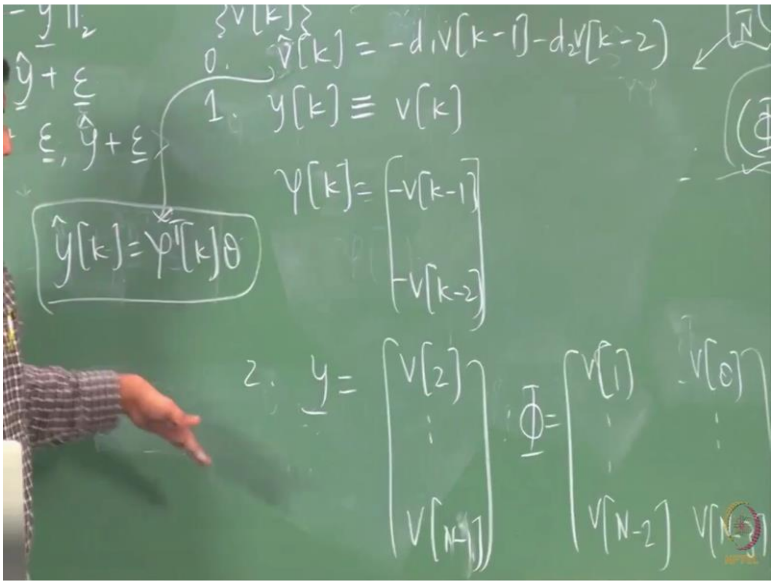
Then we will impose conditions on z we say; let us assume z to be some random signal random process and we are also making a big assumption which is that the psi that we have in this DGP is the same set of regressors that I am using; same set in the sense that I have not missed out any regressor, I will give you a simple example suppose I use least squares methods for fitting a r models, suppose I do that what would be you have to tell me what the regressor are and what y is and what is at the regressor set.

Suppose I am fitting an AR 2 model.

Student: (Refer Time: 04:18).

Y would be simply your process v k what would your; this is the place where generally students have difficulty.

(Refer Slide Time: 04:28)




In setting up the least square problem, so I given a times series v k and I am suppose to fit an AR 2 model using least squares method using the generic solution that is available to me first step is to assign y and your psi that regressors. So, y k is nothing but your v k what about the regressor.

Student: (Refer Time: 04:57).

Why do we do that in fact, even be able to write this first step, in fact, the zeroth step I should say is to write your prediction equation first write your predictor because that is what we began in least squares as well we said let y had be psi, psi k times theta or psi transpose k times theta. So, here write the prediction equation for the AR 2. So, you have here minus d 1 v k minus 1 times minus minus d 2 times v k minus 2, this is my predictor I know that already for AR 2. Now I can compare this equation with this equation here y hat of k. So, I compare this equation with the equation that I have used in linear regression from where I can see y is v and then what about the regressor vector, what is a size of it; regression vector not the matrix 2 cross one.

What are the elements, if you assume minus d 1 to be theta that is ok, but if you assume. So, you have to define what is theta first; if you say theta is d 1 and d 2 transpose then, but now comes the slightly tricky part when you are setting up the regressor matrix phi assume that I have n observations; how would your phi be.

Student: (Refer Time: 06:51).

What is a size of phi in general n by p, what is p in this case 2 good, in a generic least squares problem, when we are generic least square problem; there was no notion of lagged variables; all regressors where assumed to be available at the same times as y, but here not all regressors are available at the same time as v. In fact, when I fit at k equals 0 and a position at k equals 0, the regressors are not available. So, you have to start setting up the equations at the time when the regressors are available. So, only you can start setting up at k equals 2, as a result I throw away to observations; that means, my y vector now will start from 2 and go up to n; n minus 1 if my counting since my counting is from 0.

Likewise what is what will be your phi what is a first column of phi; phi would be v 2 let it be here v 1 up to n n minus 2 and this would be v 0 running up to n minus 3. So, you should make sure that the number of rows are the same; that is what have to watch out for when you are applying the generic formulation to a particular problem, this is how it is set up even. So, you have for example, a r dot over less that is the routine that estimates the auto regressive models for you in r using the least squares ordinary least squares method.

When you supply the data, it actually sets up this matrix and this vector and then uses the least square solution. Now this data generating process that we have given here when I said it is similar to your model, what is it in what sense it is similar that you have not missed out any regressor. So, suppose that data generating process indeed is AR 2 and you are fitting an AR 2 model then structurally they are similar that is a first thing that you have to ensure.

Suppose the data generating process for the same data generating process AR 2, you are fitting AR 1; you do not know let us say you just fitting AR 1, then structurally they are not similar. Why this is very important is because some of the results that we are going to learn for example, bias variance and so on. Assume that you have not missed out anything structurally in your regressors. Now you have to ask me in practice I said earlier; I never know the data generating model, how do I make sure that that condition is satisfied, how do I make sure that I have not missed out in my model any you know that my model is structurally coat on coat similar to the process, how do I make sure that.

Student: (Refer Time: 10:22).

Well for if I take AR 2 and AR 1; one of the regressors the regresssor that I have included significant, but I missed out one regressor. So, let me ask you very simple suppose the series has come out of AR 2, you fit AR 1 how do you figure out that your model is has a short coming.

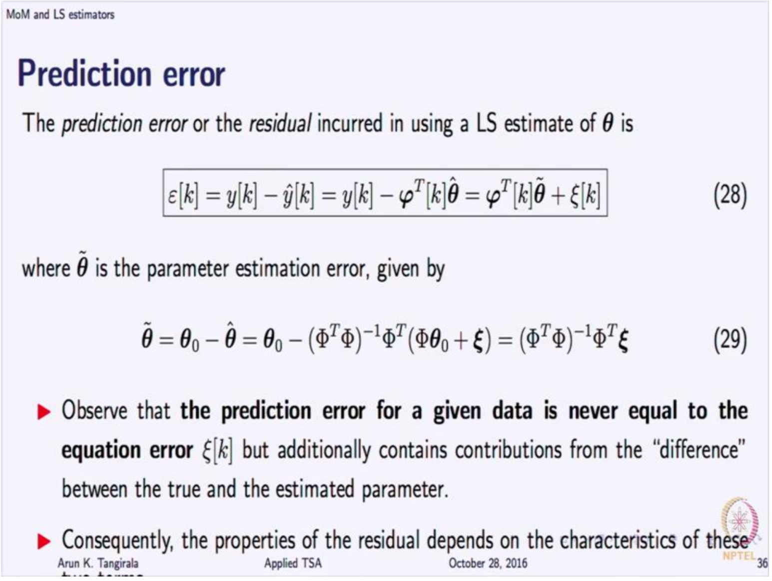
Student: Less than (Refer Time: 10:44).

Look at the residuals, so the residual analysis is the first step in any modelling exercise; to make sure that you have not missed out any regressors, in system identification there is another step in time series modelling fortunately residual analysis alone is sufficient. In system identification, you would also look at what is known as cross correlation between residual and the input, but in time series you do not have to do that just have to make sure that your residuals do not have any thing in them that you have missed out in your model.

Now, only when you have done that you can actually speak of this data generating process. So, that is one part; the other part is here you have the z k what is z k representing, what that cannot be explained by the regressors. So, it says that think of y being made up of two parts; one part that is explained by regressors and the other part that is not being explained by the regressors which is your z k. How do you know what is what it depends on the model that you assume, it simply depends on the model that you assume. Your model will define what are regressors and what is the rest, these things will become clearer later on, we only talk about the bias and then will bias is a very straight forward thing and then we will continue our discussion tomorrow.

All the properties of theta hat that we are talked about bias, variance, consistency, efficiency, distribution rest on this notion of prediction error. In fact, we have just now said residual analysis is the key, so define the residual which is y minus y hat that is very straight forward.

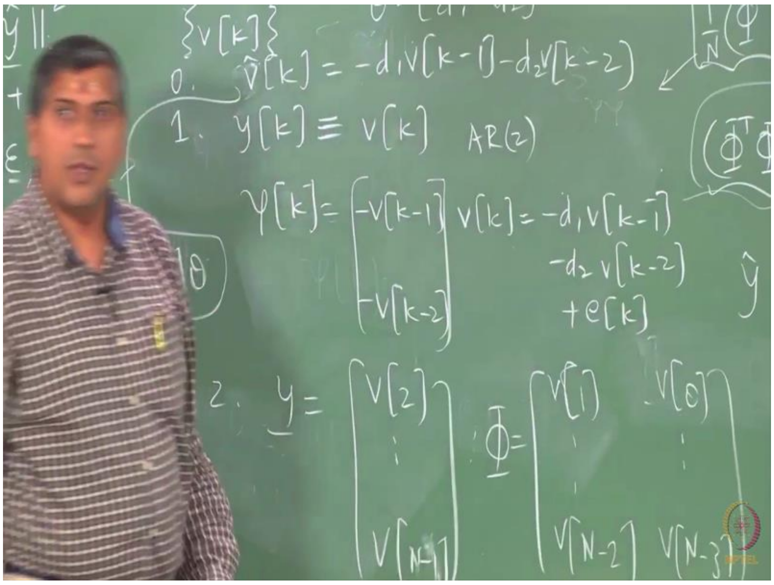
(Refer Slide Time: 12:44)




Now, notice that this epsilon if you look at equation 28; epsilon has two terms to it, what are the two terms; one, so theta naught is a truth which we do not know, theta hat is estimate that we obtained and theta tilde is the error between theta naught and theta hat, yes it is true we do not know theta tilde and reality it is only theoretical analysis.

So, you can see epsilon is made up of two terms; one coming as a contribution from the error in your estimate and the other coming from your z k which is in your DGP. Always your residual will have these two terms thus that psi transpose k theta tilde that you have is a representative of your modelling error, what you have not manage to model and then your parameter estimate error you can say so; plus z k z is something that you have not managed to include in your regressors at all. Suppose I take AR 2 DGP, we are just now said regressors are this; what would be z white noise.

(Refer Slide Time: 14:06)



Look at the data generating equation for AR 2; what is it for AR 2 the data generating equation is v k equals minus d 1 v k minus 1 minus d 2 v k minus 2 plus e k. If I have assumed AR 2 model then these are the regressors and this is my z, if I fit an AR 1 to the same process, what will be the regressor.

Student: (Refer Time: 14:40).

v k minus 1 or minus v k minus 1 what would be z k e k.

Student: (Refer Time: 14:50).

The other term that I left out, so you have to look at your z and v and the psi that regressor and whatever you left out based on the model that you have been given and DGP then only you can make a comment. So, it is not easy to say how good a method is you have to freeze the truth you have to so many things and then say under this condition, my estimator behaves this way.

So, here is your equation 29, now because this is what we are going to use for commenting on the bias for commenting; computing the variance for talking of consistency everything, what is theta tilde simply the error between the truth and estimate. We have just done some algebra here and we have managed to write this as pi transpose phi transpose phi inverse phi transpose z.

Now, this equation is the most important thing that we have to work with let me ask you a question and we will adjourn suppose I want unbiased estimates what should I expect of theta tilde.

Student: Expected value.

Expected value of theta tilde should be 0. So, if I apply that condition assume for now since we are solving generic least squares or let us say we fix phi, suppose I fix phi for a fixed set of regressors what does this condition of bias translate to on the requirement on z z z should be.

Student: (Refer Time: 16:30).

0 mean that is all, so very simple requirement; for a fixed set of regressors suppose phi is deterministic in the a r k it does not, but in general if phi is fixed then what does it mean expectation of z as long as that is 0, I will get unbiased estimates. So, that is what the condition is tomorrow what we will do is we will slightly extend this discussion to the stochastic phi, why should I worry about stochastic phi because in time series models, the regressors matrix consist of stochastic variables not of deterministic variables in that case when does least squares give me unbiased estimates and we will not talk about theory in any more, one question that I want to leave you with. Suppose you where you where to verify whether least squares give you unbiased estimates or not by simulation, what kind of simulation would you perform, how would you perform.

Let us say you are you pick any model let us say AR 2, suppose you are simulating an AR 2 process I will show that tomorrow in r, but I want you to think about it. Suppose I ask you does least squares method get me unbiased estimates of a r models, a r processes when I fit a r models to a r processes will I get unbiased estimates. Remember we have to specify the model and the DGP both question is asking that it says if I were to fit a r models to a r processes, let us say of the same order I know the order, how would you verify by simulation whether it gives you unbiased estimates.

What is a simple procedure (Refer Time: 18:21) 2 step procedure, what would you do, what is meant by unbiased estimates.

Student: (Refer Time: 18:36).

No across realization, so for different realizations I will get different theta hats and if I take the average of all those theta hats I should get the truth correct. So, the procedure would be for example, I simulate an AR 2 process fit an AR 2 model because they have to structurally match; then only I can talk about bias and so on. And then I repeat this for different realizations, I keep the process fixed, but I keep generating different realizationsl; each realization will give me different theta hat, when I average the theta hats respective theta hats, how many will I have two parameters, I average d 1 hat and d 2 hat across many many realizations, I should be sitting very close to the true values, we will being with that simulation tomorrow.