CH5230: System Identification

Fisher’s information and properties of estimators

Part 10

And very quickly I'm just reviewing what you see in the lectures. Basically in computing sigma theta hat that means we are asking whether the least square estimates are biased. And what is the variability how precise they are? We have to have a reference process. Otherwise, when I talk of bias what am I asking? For example, in bias what I'm asking? If the average of theta hats will converts to the truth. But you have to say, what is truth, you have to define. And that is what we are fixing here. We are saying assume the truth to be of this form. If I don't do this and I let the truth be anything then the

concept of bias itself doesn't make any sense. Agree? So, we are saying let the truth be this. What is the characteristic of the truth? One that structurally there is no mismatch between the regressors. I'm assuming structurally there is no mismatch between the regressors used in the process and used in my model. That is point number one to observe. We say that in other words, the plant or the process is contained in the model. My model is a superset. It should contain the plant. I may have more coefficients, see in my model have a more, included more FIR terms or then in the process or not. How many terms did I have in the model? How many do you remember?

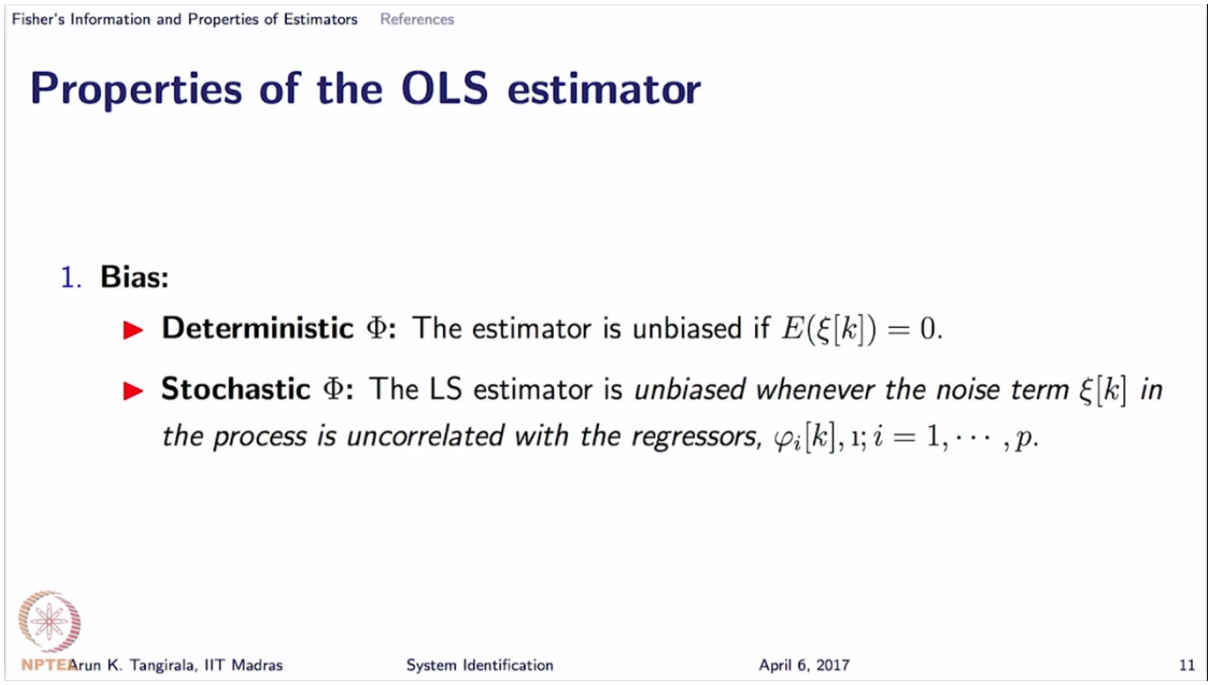
[02:04 inaudible]

Sorry. In the process, sorry. 5. k minus 5 and the delay was 2. So I had four terms. In my model I have 11 terms, right. Starting from 0 to 10. So structurally, now if you imagine the process. I can express the process. I can say that the truth theta 0 for the extra terms that I have included 0. In other words is the process contained in the model or not in our example? That's my model contain the process or not? It does. Then it's good. If it isn't, then I have to improve and that would have shown up in my cross correlation. Understand. So first, I always have to go through residual analysis before I even look at any of this bias business or variance business and so on. This is a golden thing that you should remember. Many people do not even follow this or are aware of this that the errors in the parameter estimates for any method be it least squares, Emily and so on. The expressions that are used in calculating the errors, they assume that structurally you have captured the process correctly. Why I say structurally? Because I'm saying that the regressors that have gone into generating the data should also have been used here. Yes, I have assumed a 11 coefficient FIR model. The process is a 4 coefficient FIR model. Fine, but I can still think of the process as a 11 coefficient 1 with the extra parameters being 0 value. Right? So I have included uk in my model. Is there a uk in the data generating process, no. But I can say, yes, it is there the corresponding impulse response coefficient is 0.

Right? So for my model, what would be theta 0? It would be. For my model the theta 0 would be. The first 2 elements will be 0, right. And then I have the true values and then the remaining being 0.

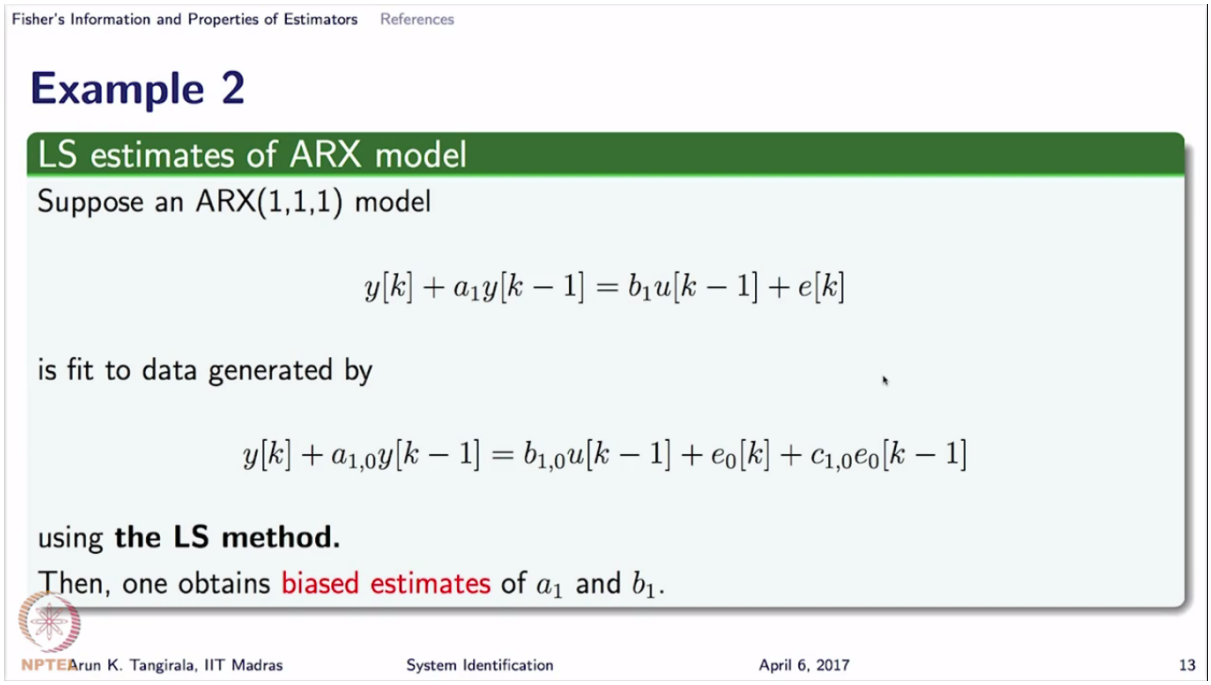
Okay. That is the hallmark of this data generating process. What is the second feature? It says that there is a randomness that is there is this other component that the regressors are not explaining, which we denote by Z. Now depending on the assumptions or the nature of Z, you will get either a consistent estimate an efficient estimate or not. That is what we're going to see. Okay. Or you will get a biased or not, unbiased estimator. So let's look at review this quickly. When you look at the bias in theta hat. So this is our theta 0 for this example, but in general, the bias in theta hat depends on two situations whether my Phi is made up of deterministic regressors or stochasticregressors. Let's worry about the stochastic part, because typically our regressors will contain some randomness also. In the FIR example it's deterministic that's okay. But very often your regressorswill contain past outputs as well. So for instance the ARX model that I'll show you shortly. Suppose I have estimating the parameters of an ARX model. What would be the regressors? Right?

(Refer Slide Time: 05:49)



Suppose I am looking at ARX 1, 1 and unit delay,the Psi would be minus yk minus 1 and uk minus 1. First order 1 coefficient, 1 delay and 1 coefficient in the numerator polynomial. ARX is only characterised by b and a. Correct. Now, the regressor contains stochasticity. Clear. So what does the result say, it says that the least squares estimator is unbiased when the Z that is whatever has been left out is uncorrelated with the regressors. Now, this is as per my model. What this result says is that, whether you will get biased estimates or not depends on that data generating process, what you have left out. So as a simple example, we'll probably skip this example. Let's look at this example. Suppose, the data is generated by this, this is the model that I assume the top is the model, the bottom is the process. Look at this carefully the top one is a model that I'm fitting and the bottom on is, the process. Now, first of all structurally are the regressors matching, right? So that part is assured that means now the problem is well post. Now you have to ask what is Z? What is Z as per this process equation? e 0 k plus c1 0 e 0 k minus 1. Correct. You should not look at the model to ascertain what is Z. You have to look at the process and say, yeah, whatever I'm leaving behind is this.

(Refer Slide Time: 07:55)



Now,what this result says is.If whatever you had left behind after, you know, beyond side transpose k theta. If that is uncorrelated with regressors you will get unbiased estimate. Now you have to tell me whether whatever you're leaving behind is it uncorrelated or correlated with the regressors?

Correlated.

Correlated, why?

Because y of k minus 1 ek minus 1.

Sorry. Because y of k minus 1 ek minus 1. Correct. So, Z is being generated by e 0 k and e 0 k minus 1, the true white-noise sequence that I'm using. And the regressor contains yk minus 1 and uk minus 1. u is not correlated with e, under what conditions?

Open loop.

Open loop conditions. Correct. And if it's close loop then there is an issue. y k minus 1 even under open loop conditions is going to be correlated with Z. So which means in this case if I fit an ARX model, if I fit an ARX model to this data, then I will obtain biased estimates. What do we mean my bias? A systematic error. So there is this MATLAB script again that's available on the webpage here.

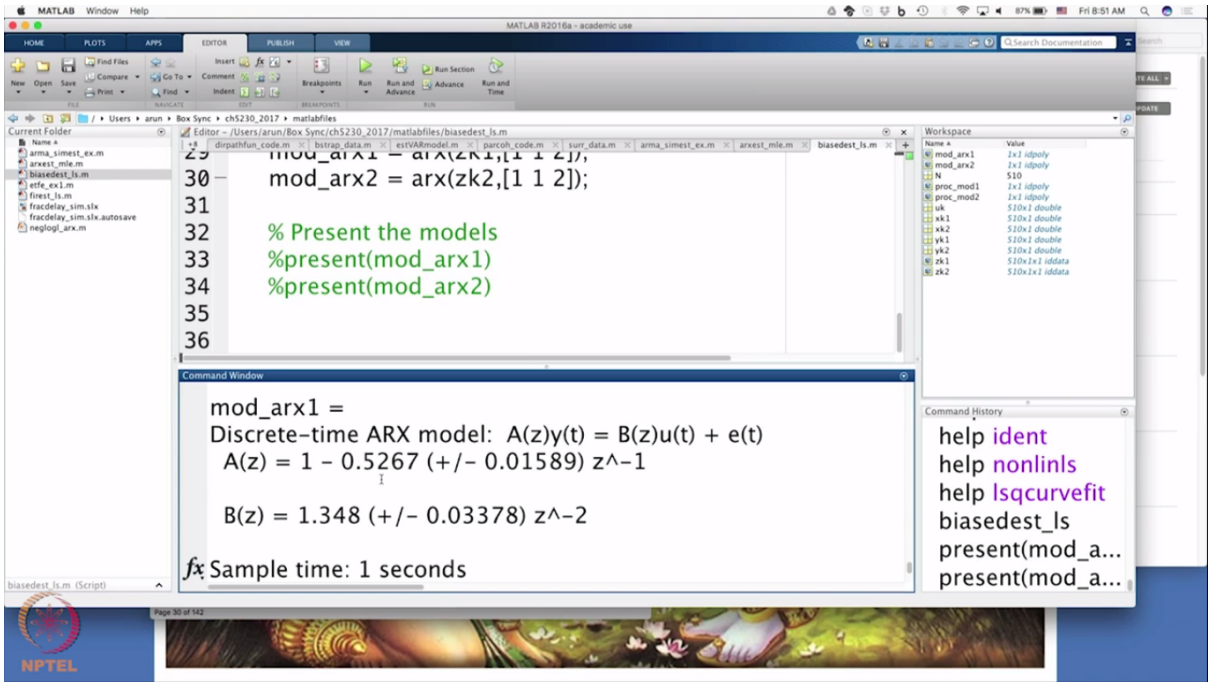
So this is MATLAB script here. I'm generating two data from two different processes. In the first process it's an ARMAX process and the second process is an ARX. All right. Now, what I'm doing is, I'm simulating and generating the data and then collecting the input output data. And now I'm using the ARX model, ARX routine to estimate, by the way this ARX routine implements least squares, linear least squares for all other cases we know for ARMAX, OE and BJ. I only get pseudo linear regression forms. I don't get the linear regression form. Whereas the ARX model alone is. And FIR models both these correspond of the linear least squares case. So, the ARX routine implements the linear least squares method. And let me see here. I'm going to run this, script here. So I've run the

script. Now let'slook at the model, so the model that we have estimated. So this is the model that I estimate here. All right. In fact. Yeah. I don't know [10:54 inaudible]. That's correct. So this is the model that I estimate in the first case. Okay. And what about the. So what are the truth, true values of for this first case. I have minus 0.5, I get minus 0.53 roughly. And then in place of 1.4 I get 1.35. And what are these? These are actually not. Let me show here. These are point estimates, but what you should do in order to check for bias. So in this case now, the first, the process one is an ARMAX process and as per our discussion, I should expect to see biased estimates. What about the second one?

So here are the estimates of the second model that is now, this is the same ARX model that's being fit to the second data set. The second data set falls out of an ARX process. If the data generating process is ARX, then what is z? Simply, white noise. Right. If the data generating process is ARMAX as perour example that we just saw on the screen in the slides, then z is not e k. That doesn't matter. The question is not whether z k is white or coloured. The question is whether the z is correlated with the regressors. So for the first dataset, if the situation is that z is actually correlated with the regressors and therefore I will see biased estimates but I cannot conclude by simply looking at the point estimates here that yes, there is a bias.

What should I do to verify if indeed there is a bias? Actually, if you look at the second process, it's very close to the truth, 0.50 something, 1.394, whereas here you have 0.53 and you have 1.35 giving you a hint that there is a bias. But I have already shown you, how do you check for bias in simulation? Sorry. So you repeat, what you should do is take the script, put it in a For loop, run it for different realisations, take the average of the corresponding parameter estimates and compare the truth. They should not be-- The difference should be negligible.

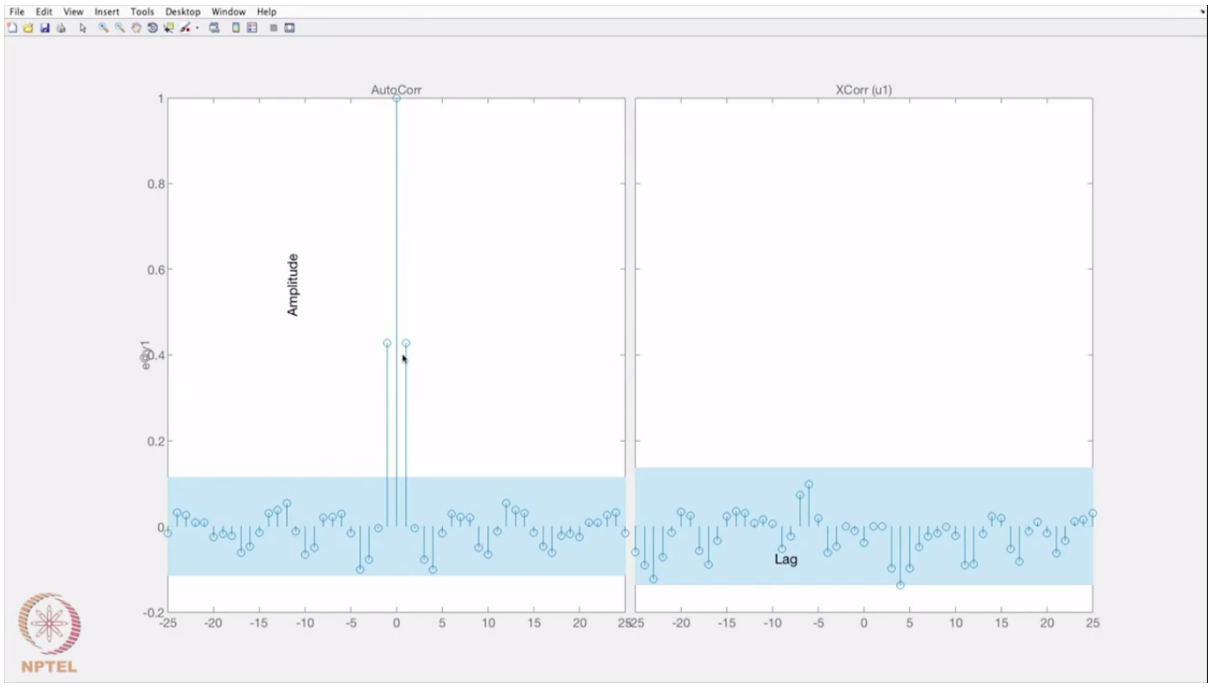
(Refer Slide Time: 13:39)



I mean if you look at percentage error, it should be extremely small. What you will find is that, that is not the case. So on one hand we have gone through the theory another hand we have the practice. So the question that you should ask is now. Fine, you have told me that whatever I have left out is uncorrelated with a regressors. I am assured of unbiased estimates but how do I know that whatever I have left out is uncorrelated with the regressors. How do I know? Cross correlation. That is why it's important to go through the residual analysis.

In your residual analysis what are you doing? You are computing the cross correlation between what you have left behind and regressors. If you find correlation then that means you already are working with a biased estimate. So what should I do? In this case for the data set1, which is coming out of ARMAX process, if I compute the cross correlation, you should do that. You can use the resid command which is built into SysID tool box. You supply the model and the data. Right. You will find that the cross correlation is significant. What do you do next? What do you do? So here is your outcome of your resid command.

(Refer Slide Time: 15:27)



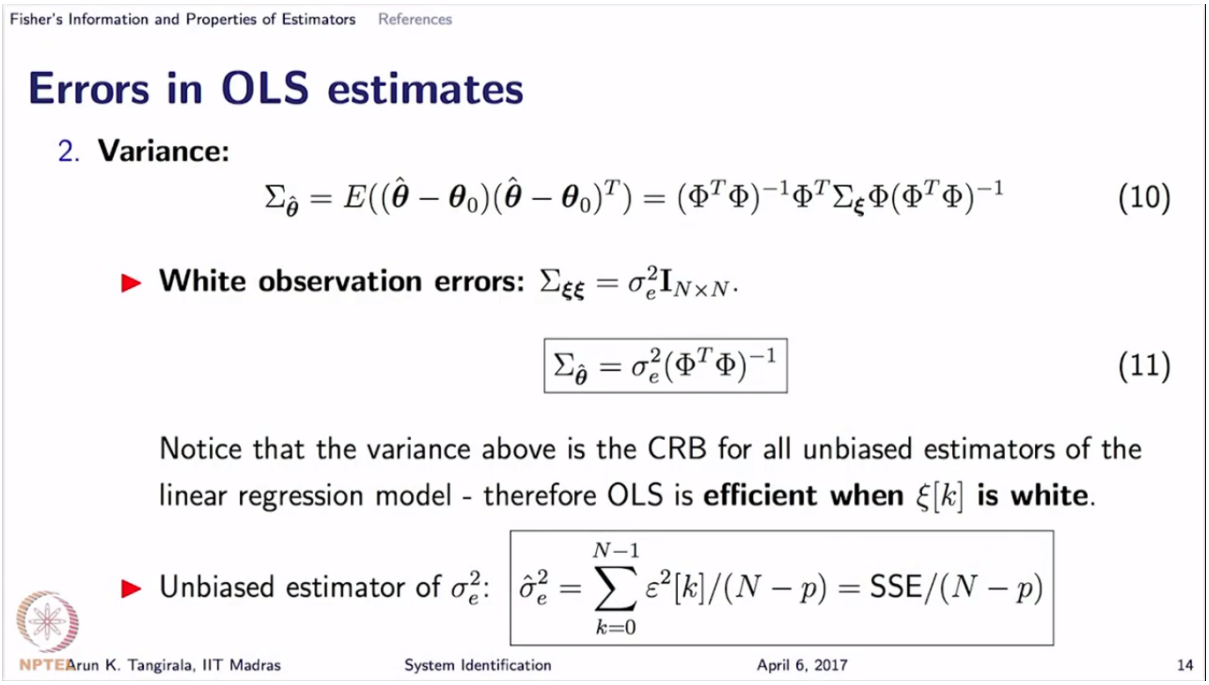
On the left hand side is a cross correlation plot, on the right hand side you have the auto correlation. Let's ignore the auto correlation for now. The cross correlation clearly tells me that whatever I have left behind is correlated with the regressors. Right. What do I do know? What is the next logical step? Quickly. Why there is silence in the room? What do I do next? What happened? I found that, whatever is left behind has some correlation with the regressors. What are the options that I have?

Sorry.

So, include, exactly, you fit higher order ARX models, you include more regressors bcause you know some effects have been left behind. But should I be actually doing that. Because look at the data generating process. It is ARMAX. What does it say? The G is first order. It is because there is a mismatch between the noise that you have assumed and what has gone into generating the data that is forcing you now to fit a higher ARX model. So if you're adamant, if you insist on fitting an ARX model then this result says, the order of the ARX model is insufficient. You have to go and improve.

What is the other option? Other option is to stick to the order of G and play around with H. And the third option we learn later on is to use another method. If you insist that I will still use an ARX model and I still want unbiased estimates, regardless of how the data was generated then you can use another method called instrumental variable method, IV method, which we'll discuss next week. Okay. The IV method will get me unbiased estimates despite the fact that the regressors are correlated with the residuals. That's the beauty of the IV method. But hopefully now, you're getting a feel of what is involved in model building, right.

(Refer Slide Time: 18:00)

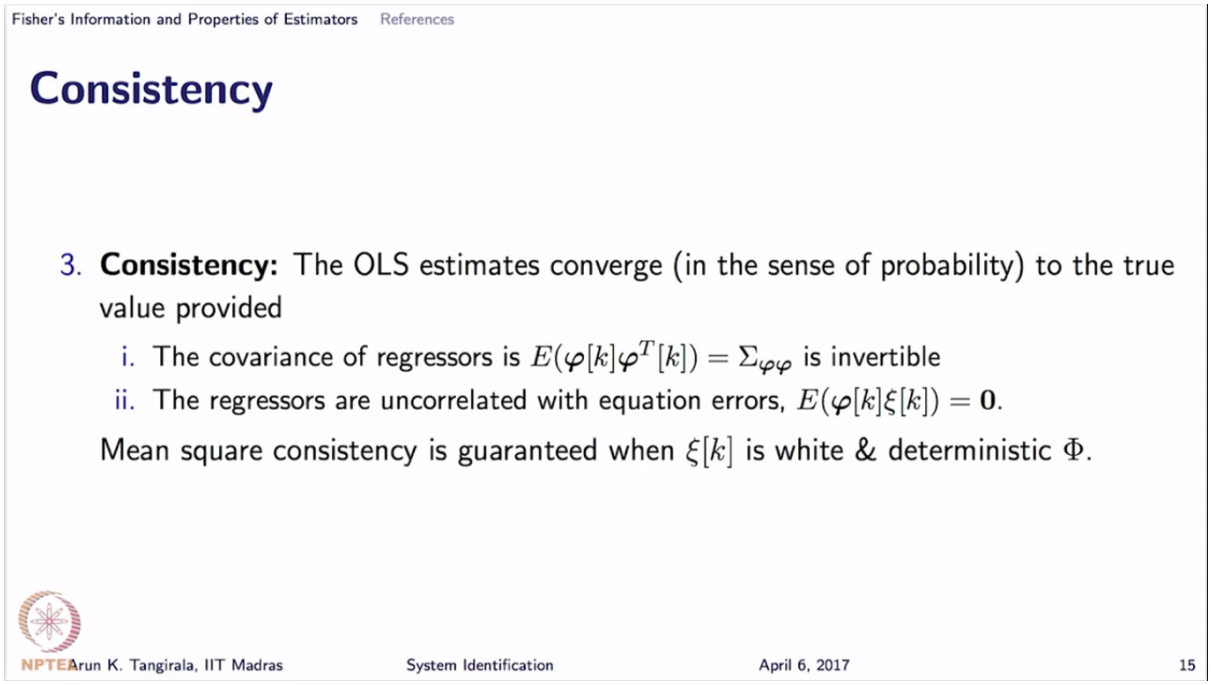


So very quickly, let me summarize things here for you in the least squares case. So we have talked about bias variance expressions generally are calculated for very specific scenario, the scenario is that your Z's are white.

In practice you should always remember, I do not know Z. What do I have to ensure therefore, I have to ensure through a careful modelling that the residuals are white. The residuals are representatives of Z. They are not equal to Z. Residuals are Y minus Y hat. Z is the actual one that has gone into the generating the data but I don't have access to that Z. I make sure that these expressions can be used by first ensuring that the residual are white. So the procedure is, if you want to use these expressions for calculating sigma, theta, hat from where you calculate the respective errors in individual theta hats, you have to ensure that the residuals are white.

And of course, you have to ensure that the residuals are uncorrelated with the regressors to guarantee unbiased estimates. Which means the residual analysis is a must. And in this expression for calculating sigma, theta, hat I need to know sigma square e, and that sigma square e is again computed from residuals. So you see the residual holds the fort, it holds the key to your fort. That's extremely important. So therefore, you should be comfortable with residual analysis.

(Refer Slide Time: 19:44)



Consistency of least square estimates, again this is just a recap of what you see in the lectures, is guaranteed when the covariance of regresses of course is invertible. Which means you're regressor matrix is not collinear or ranked deficient. And that the regressors are uncorrelated with the errors. In other words, the model should pass the residual test. And then we consistency is guaranteed.

Now, distribution of theta hat from where we construct confidence intervals and so on. You have the standard result that when the regressors are well behaved, what we mean by well-behaved is, there is no outlier and so on in the data. And that the Phi's of full rank both, then we can say that theta hat follows a Gaussian distribution asymptotically. Remember here, when we talk of theta, we are talking collectively, the parameters estimates have a joint Gaussian distribution. We are not talking of individual parameters. And what does this mean? Theta not is the mean of theta hat. That means it's an unbiased estimate. And this is a variance of sigma theta hat.

In practice, we do not know sigma square e but you estimate using the expression I gave earlier.This sigmaPsiPsi is the variance covariance matrix of the regresses which again we estimate from the data.

(Refer Slide Time: 21:13)

Distribution of OLS estimates

Theorem

If $\xi[k]$ are **independent** *and identically distributed (i.i.d.) with mean zero and variance σ^2 and the regressors are "well-behaved", then*

$$\hat{\boldsymbol{\theta}} \xrightarrow{d} \mathcal{N}\left(\boldsymbol{\theta}_0, \frac{\sigma^2}{N}\Sigma_{\varphi\varphi}^{-1}\right) \tag{12}$$

- By well-behaved regressors it is meant that
  - (i) $\Phi^T\Phi$ is of full rank as $N \to \infty$
  - (ii) No single observation shall dominate the data.
- In practice, the distribution properties are computed by replacing the theoretical quantities with their corresponding sample versions,

What we do is, once you construct the sigma theta hat. Your sigma theta hat is going to be a matrix from where we pluck the diagonal elements and construct approximate confidence intervals. Okay. So go to the lectures it'll tell you all of that but you should be, what you should be aware is that these expressions for computing confidence regions, standard errors and so on, should be used only after you have ascertained you've gone through the residual analysis.

(Refer Slide Time: 21:47)

Confidence Intervals from LS estimates

- The $100(1-\alpha)\%$ confidence interval for θ_{i0} is,

$$\hat{\theta}_i - t_{1-\alpha/2,N-p}\sqrt{\hat{\sigma}_e^2 S_{ii}} \leq \theta_{i0} \leq \hat{\theta}_i + t_{1-\alpha/2,N-p}\sqrt{\hat{\sigma}_e^2 S_{ii}} \tag{14}$$

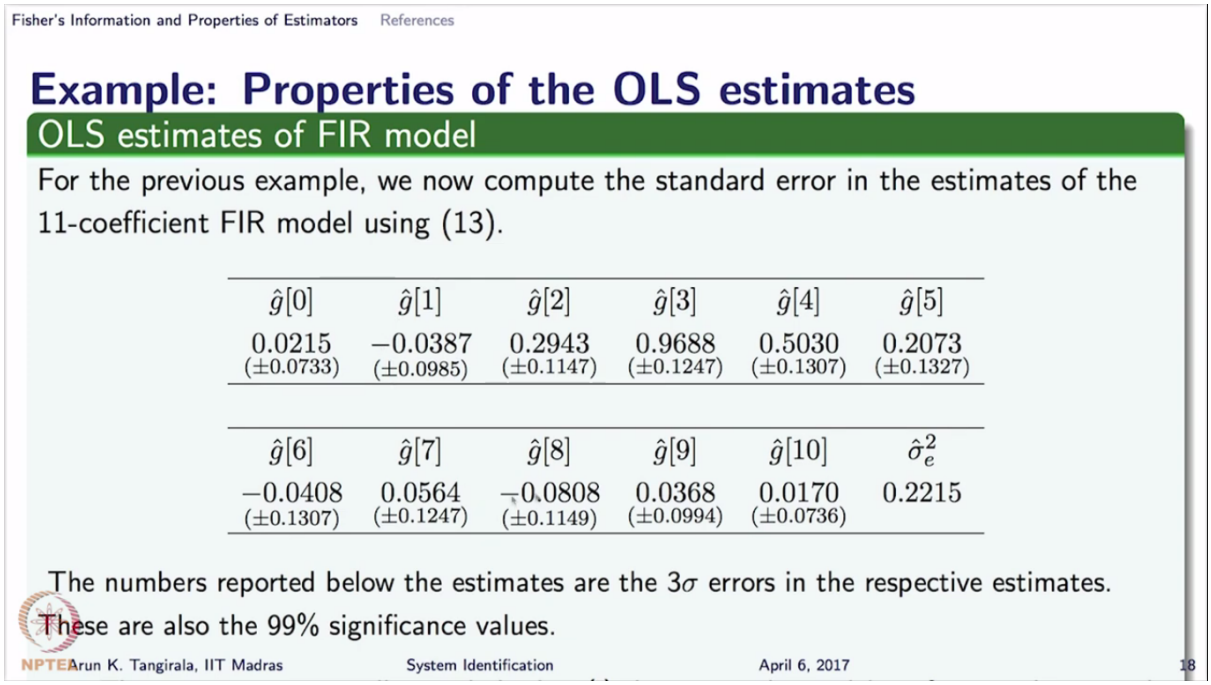
  where $t_{1-\alpha/2,N-p}$ is the critical value of the t-distribution with $(N-p)$ d.o.f.
- For large N, t-distribution tends to a Gaussian. The 99% C.I. for θ_{i0} is approximately

$$\hat{\theta}_i - 2.58\sqrt{\hat{\sigma}_e^2 S_{ii}} \leq \theta_{i0} \leq \hat{\theta}_i + 2.58\sqrt{\hat{\sigma}_e^2 S_{ii}} \tag{15}$$

Only when you are assured that there is no correlation between the residuals and the regressors and that the residuals are white, you should only then you should use these expressions. Otherwise, you should not even look at the errors. Okay. So, for the FIR model estimation problem, we have assured all of that and we have here the estimates and the three sigma errors. When you look at these errors

you will realize that apart from 2,3,4,5 the rest of the estimates, if you construct the confidence regions they will include a 0.

(Refer Slide Time: 22:25)



So take for example g0, right. g hat of 0. What you are given is three sigma here. 99% confidence region can simply be obtained by taking the point estimate and simply doing a plus or minus. Normally I report the sigma but in this case I am reporting three sigma. So 0 is obviously going to be included in the confidence region. As a result the null hypothesis that Z 0 is 0 is not rejected. And you apply this to all the coefficients, you will find that only 2,3,4,5, g's are significant and then you re-estimate and the re-estimated values are given here.

(Refer Slide Time: 23:01)

Example: OLS estimates of FIR model ...contd.

Based on the error analysis we recognize that the IR estimates are significant only at lags $l = 2, 3, 4, 5$. As a follow-up, it is necessary to re-estimate the FIR model containing only these terms. The re-estimated model is,

$\hat{g}[2]$	$\hat{g}[3]$	$\hat{g}[4]$	$\hat{g}[5]$	$\hat{\sigma^2}_e$
0.272 (±0.0692)	0.9756 (±0.0836)	0.5136 (±0.0838)	0.2073 (±0.0695)	0.2213

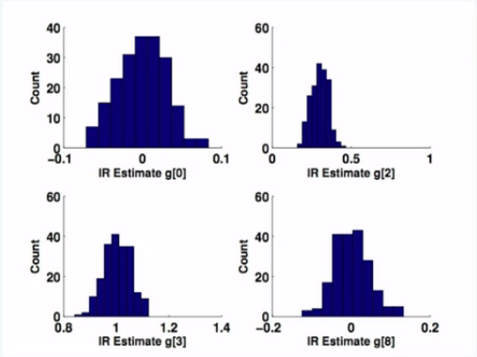
- Observe that the 3σ errors are considerably lower than those of the previously estimated 11-coefficient model. The innovations variance $\hat{\sigma}_e^2$ is also now lower.
- The improvement is due to the decrease in $\dim(\boldsymbol{\theta})$, which increases the d.o.f. for estimation.

Remember I said, after you have discovered that some of these parameters are not supposed to be included in the model. You have to re-estimate the plot model by throwing away those parameters, by omitting those parameters.

(Refer Slide Time: 23:16)

Example: OLS estimates ...contd.

To verify the distributional properties, $N_r = 200$ realizations of data are generated. The histograms of IR estimates at lags $l = 0, 2, 3, 8$ are plotted. It is fairly apparent that the estimates follow a Gaussian distribution.



And the distribution of the parameter estimates are shown to you obtain from simulations. I've done Monte Carlo simulations. Again there is a MATLAB script, you should go through that. When we meet next week, we'll conclude our discussion on least squares MLE and we'll go through a discussion on prediction error methods and then take it on from there. I will also give you a quiz papers next week. Okay. Thank you.