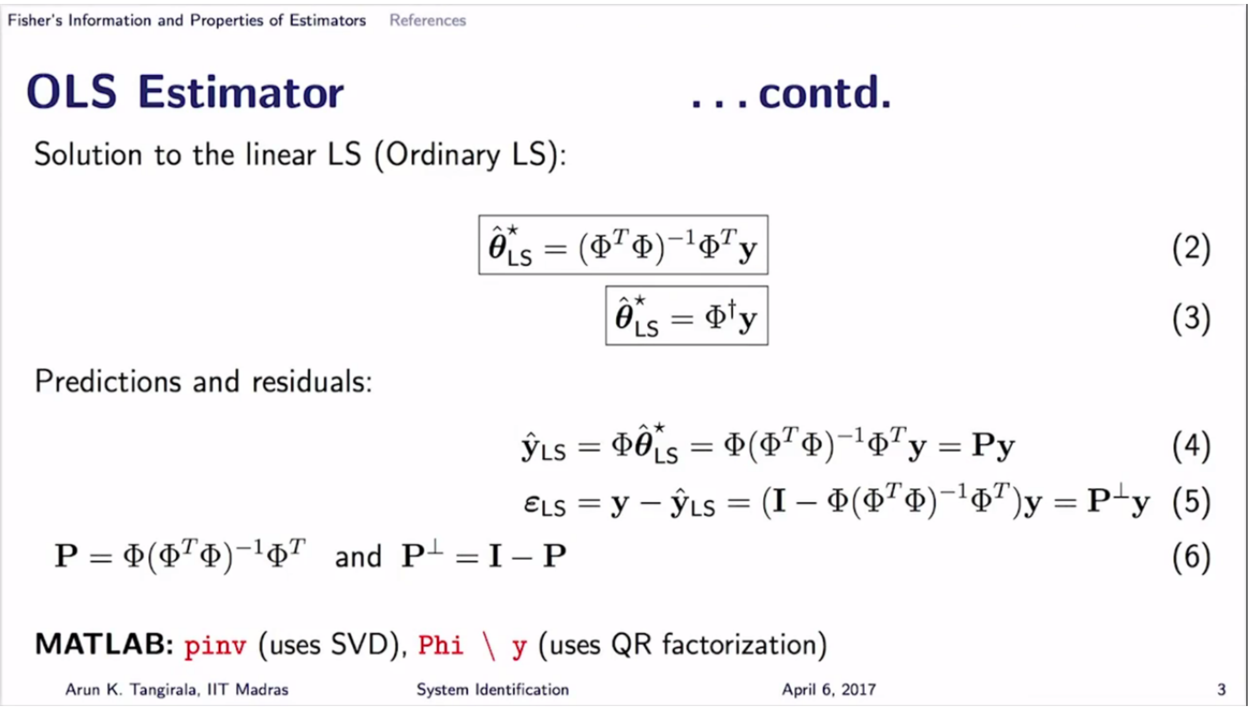
CH5230: System Identification

Fisher's information and properties of estimators

Part 09

Very good morning.

(Refer Slide Time: 0:18)



Fisher's Information and Properties of Estimators References

OLS Estimator ...contd.

Solution to the linear LS (Ordinary LS):

$$\boxed{\hat{\boldsymbol{\theta}}^{\star}_{\text{LS}} = (\Phi^T\Phi)^{-1}\Phi^T\mathbf{y}} \quad (2)$$

$$\boxed{\hat{\boldsymbol{\theta}}^{\star}_{\text{LS}} = \Phi^{\dagger}\mathbf{y}} \quad (3)$$

Predictions and residuals:

$$\hat{\mathbf{y}}_{\text{LS}} = \Phi\hat{\boldsymbol{\theta}}^{\star}_{\text{LS}} = \Phi(\Phi^T\Phi)^{-1}\Phi^T\mathbf{y} = \mathbf{P}\mathbf{y} \quad (4)$$

$$\boldsymbol{\varepsilon}_{\text{LS}} = \mathbf{y} - \hat{\mathbf{y}}_{\text{LS}} = (\mathbf{I} - \Phi(\Phi^T\Phi)^{-1}\Phi^T)\mathbf{y} = \mathbf{P}^{\perp}\mathbf{y} \quad (5)$$

$$\mathbf{P} = \Phi(\Phi^T\Phi)^{-1}\Phi^T \quad \text{and} \quad \mathbf{P}^{\perp} = \mathbf{I} - \mathbf{P} \quad (6)$$

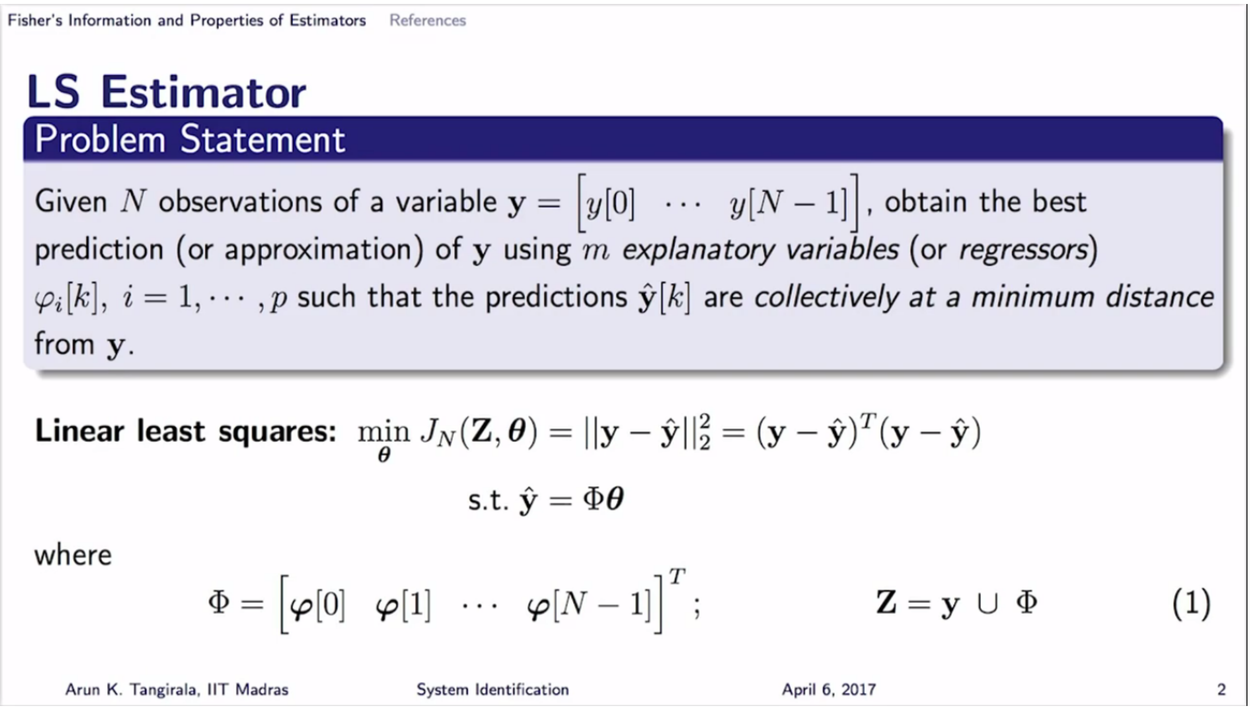
MATLAB: pinv (uses SVD), Phi \ y (uses QR factorization)

Arun K. Tangirala, IIT Madras System Identification April 6, 2017 3

What we shall do today is review the notes on least squares. And I keep saying this I hope that you have at least watch the videos on the least squares estimation. The video lectures will essentially show you how to delay the least squares estimator. Using what is known as projection theorem and goes on to discusses the properties of estimator. What we are doing is, we are learning genetic method. So whenever we discuss estimation methods be, method of moments, least squares or MLE or base N. We try to learn the genetic method. And then ask how an identification problem is cast into a particular estimation method.

Right?So here in the least squares we are assuming that. So we are assuming that there is a y hat, that means an approximation or in the context of identification we have prediction. And that predictorhas a linear form. Later on we will also ask what happens, I mean, that is what leads us to nonlinear least squares when y hat is a non-linear function of the Unknowns.

(Refer Time Slide: 01:52)



Fisher's Information and Properties of Estimators References

LS Estimator

Problem Statement

Given N observations of a variable $\mathbf{y} = \begin{bmatrix} y[0] & \cdots & y[N-1] \end{bmatrix}$, obtain the best prediction (or approximation) of $\mathbf{y}$ using m *explanatory variables* (or *regressors*) $\varphi_i[k],\ i = 1, \cdots, p$ such that the predictions $\hat{\mathbf{y}}[k]$ are *collectively at a minimum distance* from $\mathbf{y}$.

Linear least squares: $\min_{\boldsymbol{\theta}} J_N(\mathbf{Z}, \boldsymbol{\theta}) = ||\mathbf{y} - \hat{\mathbf{y}}||_2^2 = (\mathbf{y} - \hat{\mathbf{y}})^T(\mathbf{y} - \hat{\mathbf{y}})$

$$\text{s.t. } \hat{\mathbf{y}} = \Phi\boldsymbol{\theta}$$

where

$$\Phi = \begin{bmatrix} \boldsymbol{\varphi}[0] & \boldsymbol{\varphi}[1] & \cdots & \boldsymbol{\varphi}[N-1] \end{bmatrix}^T; \qquad \mathbf{Z} = \mathbf{y} \cup \Phi \qquad (1)$$

Arun K. Tangirala, IIT Madras System Identification April 6, 2017 2

So the important thing to recognize here is, we are writing y hat as Pi theta. This y hat is either an approximation or a prediction. For a particular identification problem the only thing that remains is when you take this least squares estimation to identification is to recognize what is Phi, what is theta. Whether you can write the prediction problem that the predictor in that form and if you can, then what is Phi, what is theta? We will go through a couple of examples today. I'll show you a couple of examples in MATLAB which pertains to estimating either the FIR model coefficients or ARX model coefficients.

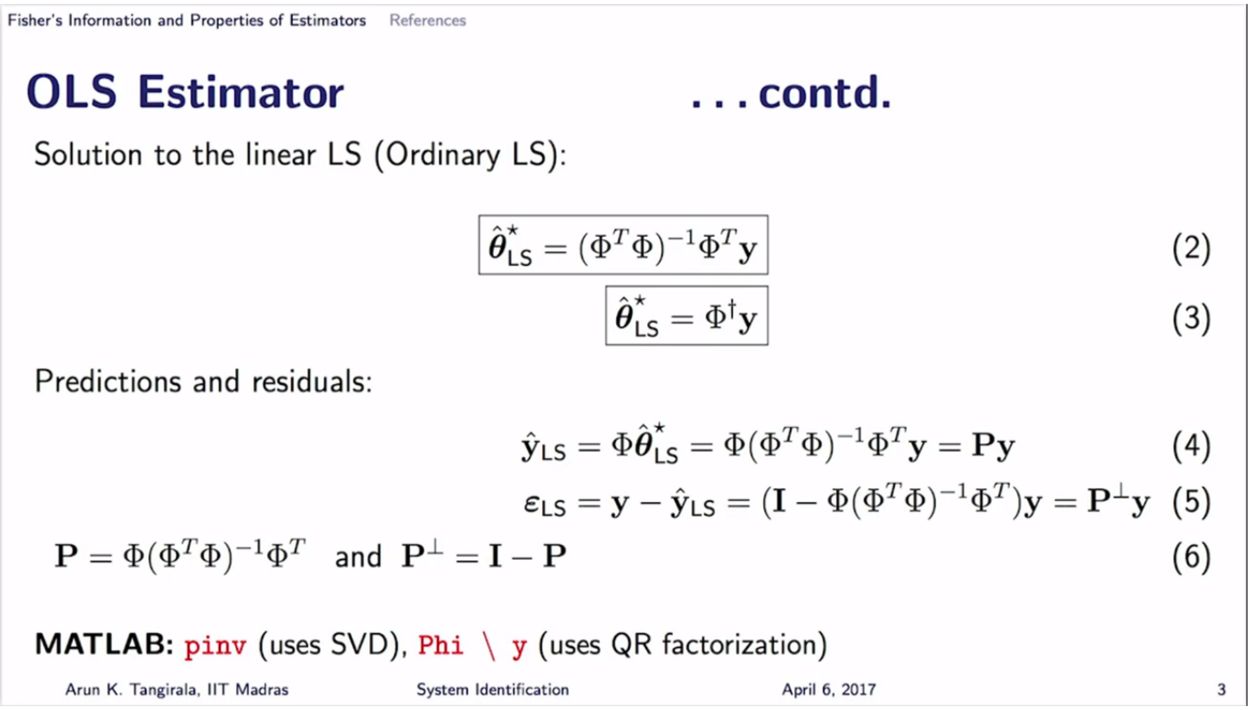
So there are two things here,one you're learning the generic method of estimation. And two, you will have to be now well versed with taking an identification problem. What is a identification problem that remains the same right from the beginning. I'm given data and I'm supposed to estimate G and H and sigma squaring. That has not changed at all and it will remain the same. How do we cast that problem into this genericframework? It's always useful to study the genetic problem because then given any other estimation problem tomorrow you can just cast that problem into the generic formulation. And then use the solution.

So the generic formulation here is y hat is Pi theta or Phi theta, big Phi. And this big Phi is a matrix in general of N observations by p. Right?Phi is N by p. The number of regressors. And this Psithat you have is a regression vector. What is the size of Psi? Sorry. No that isPhi. Sorry. p by 1. Okay. A y defaultour vectors are always column vectors. The Psi k is always p by 1. Theta is p by 1. And the big Y [04:25

inaudible]. So you should, as long as you remember this basic expression you can derive the dimensions even on the fly.

Okay. So the least squares problem is that of minimizing the sum square approximation error subject to y hat equals Phi theta. And the solution turns out to be Phi transpose Phi inverse Phitranspose y. What happened?It's okay. P cross always. y hat of k is a scalar. So Psi k is p by 1. Psi transpose k would be 1 by p. Theta is p by 1. Clear? All vectors are column vectors by default. So Psi transpose will be row vector. Okay. So here the solution theta have this Phi transpose Phi inverse Phi transpose y.

(Refer Slide Time: 05:36)



Fisher's Information and Properties of Estimators References

OLS Estimator ...contd.

Solution to the linear LS (Ordinary LS):

$$\boxed{\hat{\boldsymbol{\theta}}^{\star}_{\text{LS}} = (\Phi^T\Phi)^{-1}\Phi^T\mathbf{y}} \qquad (2)$$

$$\boxed{\hat{\boldsymbol{\theta}}^{\star}_{\text{LS}} = \Phi^{\dagger}\mathbf{y}} \qquad (3)$$

Predictions and residuals:

$$\hat{\mathbf{y}}_{\text{LS}} = \Phi\hat{\boldsymbol{\theta}}^{\star}_{\text{LS}} = \Phi(\Phi^T\Phi)^{-1}\Phi^T\mathbf{y} = \mathbf{P}\mathbf{y} \qquad (4)$$

$$\varepsilon_{\text{LS}} = \mathbf{y} - \hat{\mathbf{y}}_{\text{LS}} = (\mathbf{I} - \Phi(\Phi^T\Phi)^{-1}\Phi^T)\mathbf{y} = \mathbf{P}^{\perp}\mathbf{y} \qquad (5)$$

$$\mathbf{P} = \Phi(\Phi^T\Phi)^{-1}\Phi^T \quad \text{and} \quad \mathbf{P}^{\perp} = \mathbf{I} - \mathbf{P} \qquad (6)$$

MATLAB: `pinv` (uses SVD), `Phi \ y` (uses QR factorization)

Arun K. Tangirala, IIT Madras System Identification April 6, 2017 3

Very often we also write this as theta.I mean theta hat least squares as Phi the dagger here and this dagger is use to denote a pseudo inverse of Phi. What you are actually solving is y hat equals Phi theta. Right? You assume for example, what? I mean,loosely speaking that you force Phi theta to explain y. That is what you are actually doing. By writing y hat equals Phi theta and by minimizing the sum square two errors. In effect what you are forcing is Phi theta to explain y. But you have more number of equations than unknowns and Phi is a rectangular matrix. If Phi was a square matrix then you'll get an exact solution that would be simply theta hatwould be simplify Phi inverse y. But because Phi is rectangular. You think of a pseudo inverse and that pseudo inverse is nothing but PhitransposePhi inverse Phi transpose.

It's easy to verify that when Phi is square it reduces to the regular inverse. Now there are other implications. I talk about that in the video lectures. Hopefully you should see that and understand. Now once you obtain the parameter estimates you can compute the predictions and residuals during this expressions, that's fairly straightforward. Once you get your theta hat, plug it back into your y hat then you will get your prediction and residuals. What this expression show you is, that you do not have to actually compute theta hat. You can straight away compute be prediction without computing theta hat. If you look at the expression, right?

Because once I substitute for theta hat, everything is in terms of Phi and y. So it is possible to compute these predictions and residuals. And typically we denote, I mean, we would like to think of these predictions as some kind of projection matrix times y. This p is a projection matrix and one of the most important properties of the least square solution that you should remember is that the residuals are orthogonal to the predictions or regressors. You can see predictions also because predictions are made up of regressors. That is a unique property of least squares estimate.

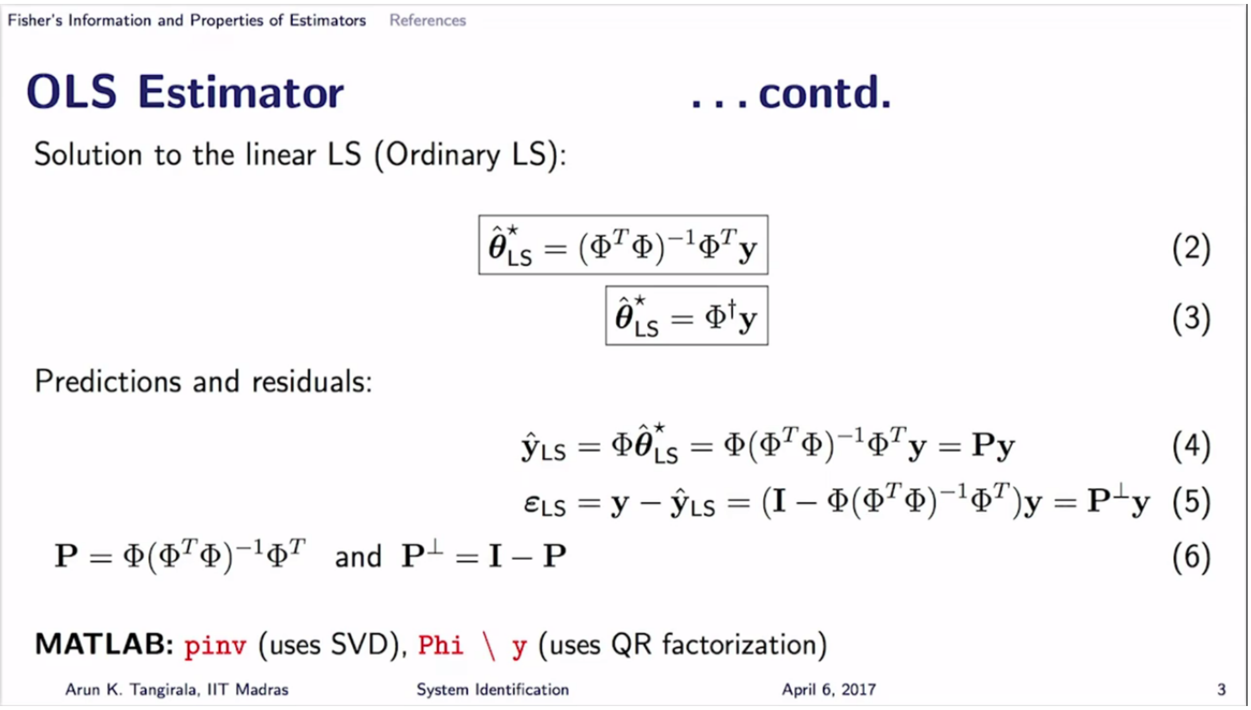
It generates approximation in such a way that the residuals are orthogonal to them. Which means there is nothing left in the residuals that you can really skim further to improve your prediction. You can do much in a linear sense that is what we mean by being orthogonal. In MATLABthe least squares estimates are obtained by either using pinv or the backslash operator. Now I have explained in the videos, what is a numerical way of implementing leastsquares.One the handwritten solution is Phi transpose Phi inverse Phi transpose y.
That's Okay. On paper that looks good. But when it comes to practice you can have different situations particularly remember you're going to invert this matrixPhi transpose Phi.

And inversion we know is computationally sensitive. Small errors in Phican really add up and also computing inverse is not such a friendly operation, computationally friendly operation. So in order to guarantee numerical robustness of this inverse computation, it's a standard practice to use QR factorization or SVD. The pinv command in MATLAB uses the SVD method of computing the least squares again either you can refer to my textbook or to the lectures to understand how SVD comes into play in computing the least squaresestimate.

And this backslash operator uses what is known as QR factorization, once again refer to the textbook or the video lectures. Both these methods of computing the least squares estimate, what they do essentially is they compute a numerically stable in inverse of Phi transpose Phi. That is a major point to keep in mind. While SVD does it in one way, the QR factorization does it in another way. See, many a times what can happen in general, remember we are studying a very genetic method. Manier times the PhitransposePhi can be singular or close to being singular. That means your Phi matrix has a lot of regressors that or has regressors that look very identical, that are collinear we say. When you have collinearity in a matrix then it becomes a rank deficient.

And then in computing the inverse becomes difficult. And that can happen when you have included way more regressors or not just way more regressors, regressors that are very closely related. You shouldn't have included them but by mistake you have done that and that leads to rank deficiency of a transpose Phi, and that leads to in turn problem with the computation of inverse of a transpose Phi. So to handle all such situations, you do not directly compute the inverse of a transpose Phi, but rather use this as SVD or QR factorization. That's all I intended to say in this talk about the implementation part.

(Refer Slide Time: 11:44)



Fisher's Information and Properties of Estimators References

OLS Estimator ...contd.

Solution to the linear LS (Ordinary LS):

$$\boxed{\hat{\boldsymbol{\theta}}^\star_{\text{LS}} = (\Phi^T\Phi)^{-1}\Phi^T\mathbf{y}} \tag{2}$$

$$\boxed{\hat{\boldsymbol{\theta}}^\star_{\text{LS}} = \Phi^\dagger\mathbf{y}} \tag{3}$$

Predictions and residuals:

$$\hat{\mathbf{y}}_{\text{LS}} = \Phi\hat{\boldsymbol{\theta}}^\star_{\text{LS}} = \Phi(\Phi^T\Phi)^{-1}\Phi^T\mathbf{y} = \mathbf{P}\mathbf{y} \tag{4}$$

$$\varepsilon_{\text{LS}} = \mathbf{y} - \hat{\mathbf{y}}_{\text{LS}} = (\mathbf{I} - \Phi(\Phi^T\Phi)^{-1}\Phi^T)\mathbf{y} = \mathbf{P}^\perp\mathbf{y} \tag{5}$$

$$\mathbf{P} = \Phi(\Phi^T\Phi)^{-1}\Phi^T \quad \text{and} \quad \mathbf{P}^\perp = \mathbf{I} - \mathbf{P} \tag{6}$$

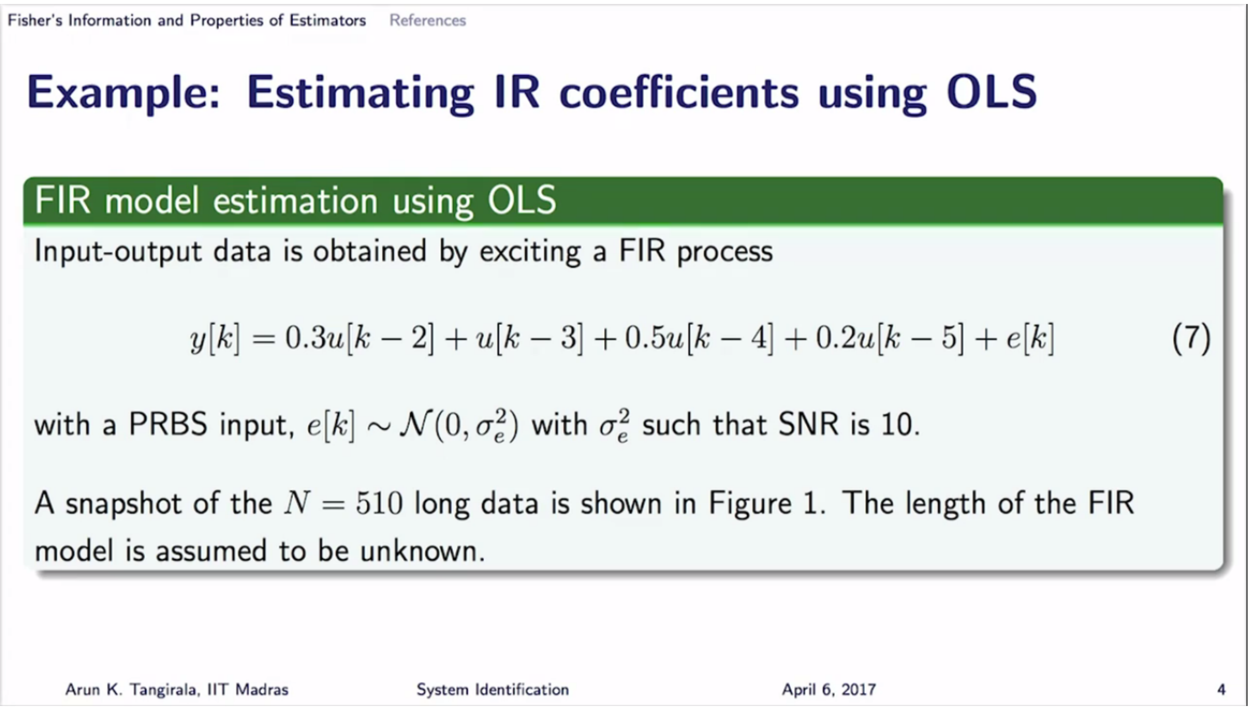
MATLAB: pinv (uses SVD), Phi \ y (uses QR factorization)

Arun K. Tangirala, IIT Madras System Identification April 6, 2017 3

But for more details you should refer to the textbook or there are much more details given in the literature on least squares implementation. But you should know that in practice it is not advisable to directly-- when you write a MATLAB code, it is not advisable to use inverse or a transpose Phi times, Phi transpose times y, not recommended. It's better to use either this Phi in or backslash. Okay, so let me actually show you an example now of how you would cast a typical identification problem. There is parameter estimation problem in SYSID, in the least squares framework.

So let's look at this example here. I have-- let us say data coming out of an FIR process. So the equation 7 that you see here is not the model yet. What is it? It is the data generating process as we call DGP. Very-- you should get used to this concept in estimation. There is a process that generates the data. There is a model that the user postulates. Equation 7 is telling you how the data is being generated. And it also says that a PRB as input is used and the variance of noise is adjusted such that that the SNR 10. Why am I giving you this information? So that if you want to simulate this data you should know, what is the generating process and so on? Now what is the delay in this process? Two, good. Now, after we generate the data how many observations are we generating about 500 and in fact, 510 observations. We assume that we neither know the delay nor the length of their FIR model.

(Refer Slide Time: 13:42)



Fisher's Information and Properties of Estimators References

Example: Estimating IR coefficients using OLS

FIR model estimation using OLS

Input-output data is obtained by exciting a FIR process

$$y[k] = 0.3u[k-2] + u[k-3] + 0.5u[k-4] + 0.2u[k-5] + e[k] \qquad (7)$$

with a PRBS input, $e[k] \sim \mathcal{N}(0, \sigma_e^2)$ with σ_e^2 such that SNR is 10.

A snapshot of the $N = 510$ long data is shown in Figure 1. The length of the FIR model is assumed to be unknown.

Arun K. Tangirala, IIT Madras System Identification April 6, 2017 4

So we begin with the standards SYSID exercise and go ahead and fit. So here is a snapshot of the input output data. Now we go ahead and fit in FIR model, of some length M. I do not know what is the length? I do not know what is I delay, it's an non-parametric model, so, I'm just beginning to learn understand the process. So just as a guess, let us estimate at 11 coefficient FIR model. It's not because I am fond of 11 or anything like that, but it is just to start off it. You could even start with 20, doesn't matter. Now, I would like to estimate a 11 coefficient FIR model. Now, note that I have written here y hat of k. If you actually see it's a predictor. It's not the y.

So what have I assume here. I've assumed that, y(k) is this expression plus e(k). I assume that the observation error it is white in writing this predictive. That's very important to remember. Now, given this y hat, now you have to tell what psi(k) is? So for this problem, what is your regressor vector. And what is theta? Should be able to tell it very quickly. This is what you should get used to. I'm sorry. Input, no you have to be much more specific, what is theta. Let us first fix theta. First of all you have to ask, can I write it in the linear least squares form. Yes or no? Yeah. That's the first thing you should establish. Now having established that what is theta? The impulsive response quite efficiently, right? So theta is-- good.

So theta is g0 running up to g10, a column vector. Very good. What about the regressor vector? What would they be?So, what is the first element? So, let's write here, so that you clear it's a column vector. This is also column vector, theta. What is the first element of Psi k? You?u k, so it starts with the u k and then the next element is k minus 1 and the last element would be. So, this is your regressor vector, right? Now, when we study the generic least squares problem or, you know, in a classical linear regression framework.

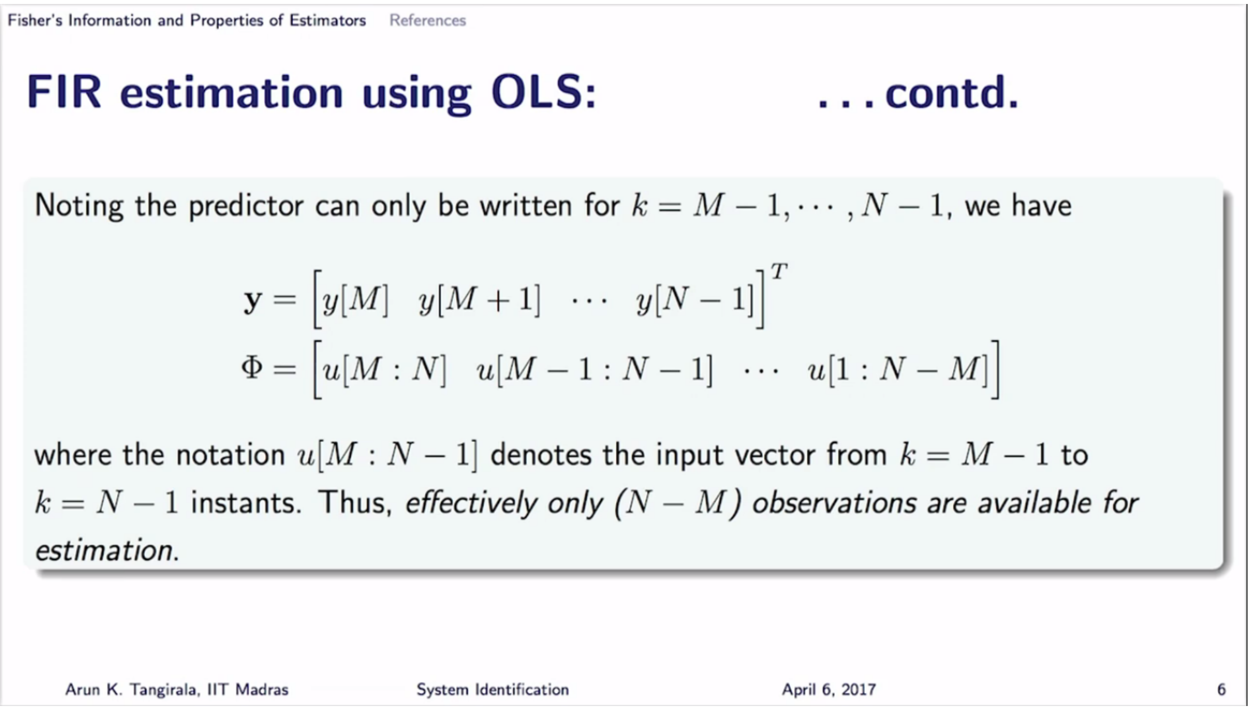
There is a difference in the sense that we assume that there are p regressors. Here also p regressors. Here, what is p here, for our problem? What is p for our problem? 11. Good. So here also, I have p regressors, but there is a difference between the generic case and this case. In the generic case, we assume that the p regressors are available at the kth instant. Right? You think of Psi 1 k, Psi 2 k, that is up to Psi p k.

We assume that all those regressors are available at-- so if you look at this equation here. This is made up of p regressors. I assume that p regressors are available at the kth instant. That is in a generic scenario. But if I look at the problem in hand, do I have the regressors available at the kth instant right from k equals 0? That means can I write the prediction expression starting from k equals 0? Or not? Yes or no? What is the answer?
Can I set up prediction expression right from k equals 0? I cannot. Because in this specific case that regressors happen to be lagged variables, the column is lagged variables of a single variable.

In a generic static linear regression problem, where the p regressors are not lagged variables. They are essentially-- what are they? They are just some p generic regressors, you know, one could be flow, other could be temperature, whatever. Right? Whatever regressors you have, they are all available at the kth instant. This is very important to note, when you are dealing with dynamic models, typically our regressors matrix will consist of lagged variables. And because you have lagged variables, you cannot set up the prediction right from k equals 0. So here, when can I start setting up the regressors vector? At what k, at what instant onwards? Really? From k equals 10.

Our counting begins from 0. Right? So I have to wait until the 10th instant to be able to set up the regressor vector. Which means although I have 510 observations, I'm not going to use all the 510, I'll have to throw away, how many observations in the beginning? How many? 10 observations I have to throw away from k equals 0 to 9, I cannot write the prediction, right? So in other words I can start rating this prediction expression for this problem, you have to understand. Only from y 10 onwards.So here I can write g 0 u 10 plus up to g 10 u 0. Only then I can start writing the prediction. y hat of 11 and so on up to y hat of 510, in fact of 509 because I have 510 observations. So what have we done now, we have thrown away 10 observations. And that's an important point to remember. We'll recall this aspect a bit later. So now my Phi therefore will begin from u M to N and so on.

(Refer Slide Time: 20:43)



Fisher's Information and Properties of Estimators References

FIR estimation using OLS: . . . contd.

Noting the predictor can only be written for $k = M-1, \cdots, N-1$, we have

$$\mathbf{y} = \begin{bmatrix} y[M] & y[M+1] & \cdots & y[N-1] \end{bmatrix}^T$$
$$\Phi = \begin{bmatrix} u[M:N] & u[M-1:N-1] & \cdots & u[1:N-M] \end{bmatrix}$$

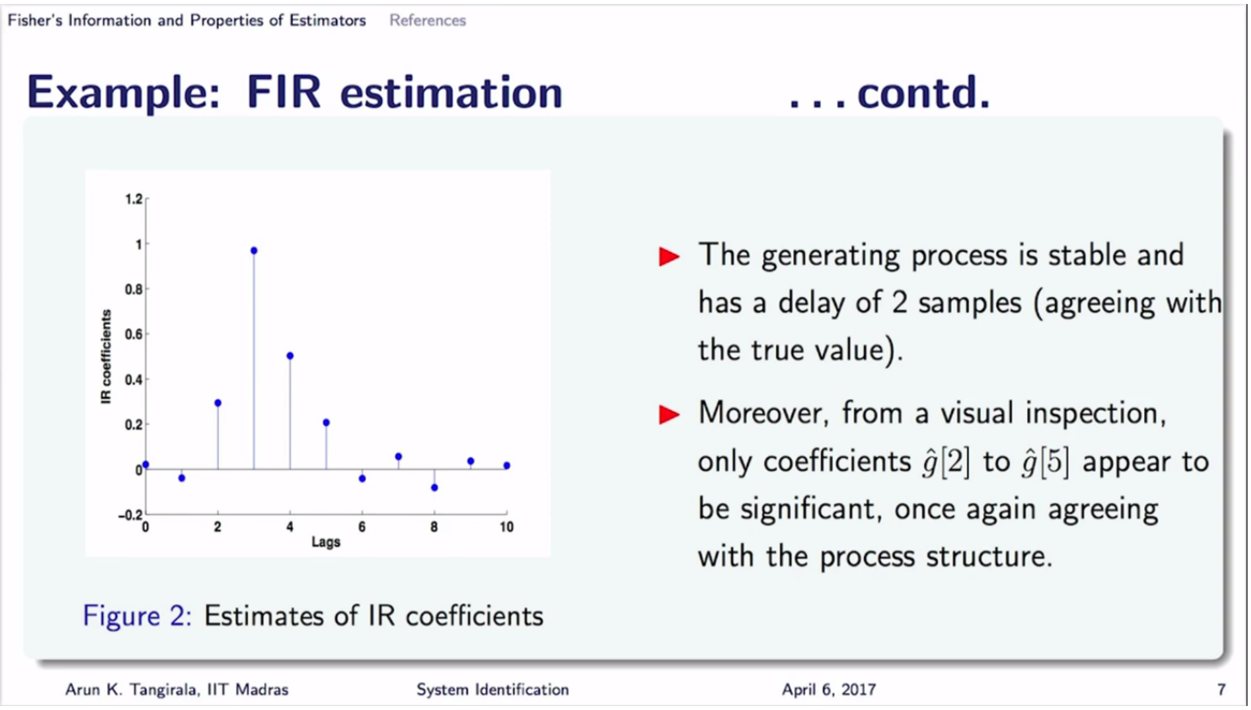
where the notation $u[M:N-1]$ denotes the input vector from $k = M-1$ to $k = N-1$ instants. Thus, *effectively only $(N-M)$ observations are available for estimation.*

Arun K. Tangirala, IIT Madras System Identification April 6, 2017 6

So that the Phi matrix is constituted accordingly, the y vector is constituted this way. So what we have done is effectively we have only N minus M observations. Correct? We have lost a few observations. This is always the case when I use least squares method or a method like this that effectively the number of observations has fallen down.

Okay.So now I'm all set. My Phi is ready, my y is ready. I'm all set to use a least squares estimator, by using the formula Phi transpose Phi inverse Phi transpose y. Of course, in MATLAB what do I do? I actually either use Phi backslash y or I simply use pin v of Phi times y. This MATLAB script, this example is straight from the textbook. I hope you already visited my website at some point in time. The scripts are available, on the web page. Please download that script and go through every line by yourself, do not copy and paste. For your own benefit, write the code. Look at code and write, that's okay. Make mistakes and then only you'll learn. But please, I do not advise you to copy and paste the script.

(Refer Slide Time: 22:07)



Fisher's Information and Properties of Estimators References

Example: FIR estimation . . . contd.

Figure 2: Estimates of IR coefficients

- The generating process is stable and has a delay of 2 samples (agreeing with the true value).
- Moreover, from a visual inspection, only coefficients $\hat{g}[2]$ to $\hat{g}[5]$ appear to be significant, once again agreeing with the process structure.

Arun K. Tangirala, IIT Madras System Identification April 6, 2017 7

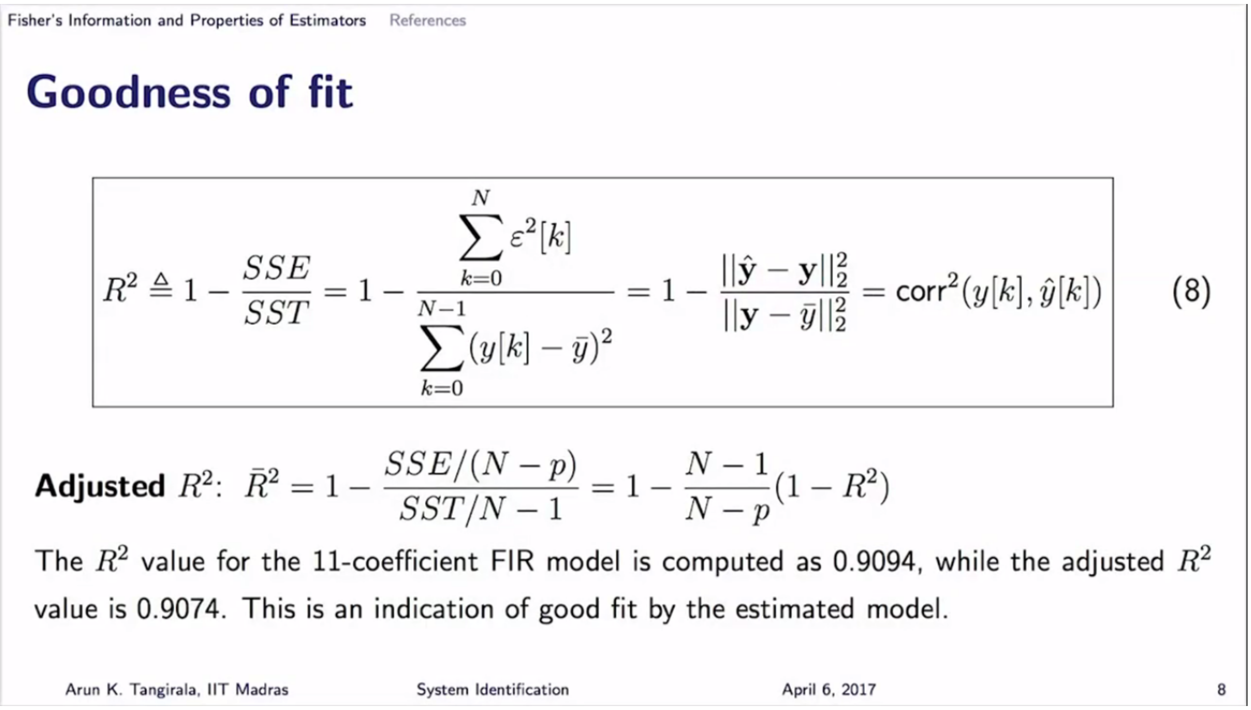
Okay, so I run the script and I generate these estimates. All right? Now when you look at these estimates the first inference that you draw is that the process is stable, right? What else can we say? We can look at these coefficients and we can say, yeah. Okay. Now, how many estimates look fairly significant? The first two look very tiny. The 2, 3, 4 and 5 look fairly significant. I'm not saying this because I know the process. I am only basing my inference on the estimates. Pretend, now we're pretending, we don't know anything about the process. The remaining ones are small. Correct? But of course, if what I call a small may not be agreeable to you.

So what is the best way to use a statistical approach, to be able to say, which of these estimates are insignificant or significant. That means now I have to run a significance test on each of these coefficients. And at each lag I'm going to test this hypothesis. Remember I spoke about this yesterday as well. So at each lag we want to test this hypothesis. So what I need is a significance level. What we mean by significance level, is very straightforward. We ask if the truth is 0, I know, when the truth is 0, the estimates are not going to be 0.

Right? Just because the true value is 0, it doesn't mean that the estimate is going to be 0. When we talk of significance levels, typically we talk of 95% significance levels, 99% significance levels and so on. What is the meaning of that? The meaning is, if the truth is 0, what is the 95% probability region for the estimates? For any parameter, for any parameter when you talk of significance level, this is the interpretation. When the true value is 0, we know estimates are not going to be 0 valued, but then what is that 95% of the times, what is a band in which the estimates will fall?

So that I can draw that band and then see if the estimate is within that band. If the estimate is within that 95% significance band, then we say that the truth is, the null hypothesis holds. If the estimate falls out of that band then we reject the null hypothesis. So it is essentially closely related to confidence intervals, although the confidence interval is slightly different from significance level, but they related. So how do we draw those significance levels? For that we need to derive the distributions of theta hat from least squares.

(Refer Slide Time: 25:24)



Fisher's Information and Properties of Estimators References

Goodness of fit

$$\boxed{R^2 \triangleq 1 - \frac{SSE}{SST} = 1 - \frac{\displaystyle\sum_{k=0}^{N} \varepsilon^2[k]}{\displaystyle\sum_{k=0}^{N-1} (y[k] - \bar{y})^2} = 1 - \frac{||\hat{\mathbf{y}} - \mathbf{y}||_2^2}{||\mathbf{y} - \bar{y}||_2^2} = \text{corr}^2(y[k], \hat{y}[k])} \qquad (8)$$

Adjusted R^2: $\bar{R}^2 = 1 - \dfrac{SSE/(N-p)}{SST/N-1} = 1 - \dfrac{N-1}{N-p}(1 - R^2)$

The R^2 value for the 11-coefficient FIR model is computed as 0.9094, while the adjusted R^2 value is 0.9074. This is an indication of good fit by the estimated model.

Arun K. Tangirala, IIT Madras System Identification April 6, 2017 8

So that is one thing that we need to do. The other thing that generally is done in least squares estimates, we'll come back to that example. Let's, so we're just taking a 2 mean D2 and a review of the goodness of least squares fits. The other thing that is normally reported in least squares methods is this R square measure. Telling you how good the fit is, it's not talking about theta hat. It is just talking about y hat. There is a difference between talking about y hat and theta hat. R square looks at how good is y hat compared to y.

Have I predicted, very well, first, because the first job is prediction. Once I'm assured that the y hat matches y very well, then I'm going to ask the question, has over fitting occur. I'm going to derive the errors in the parameter estimates and so on. But first I want to make sure that in my regressor matrix I have chosen enough regressors, that is number one,and that the least squares method has done a good job of constructing this optimal y hat. This R square measure is very popular. It's essentially a correlation between in fact, squared correlation between y k and y hat k, there is a mistake, it's not.

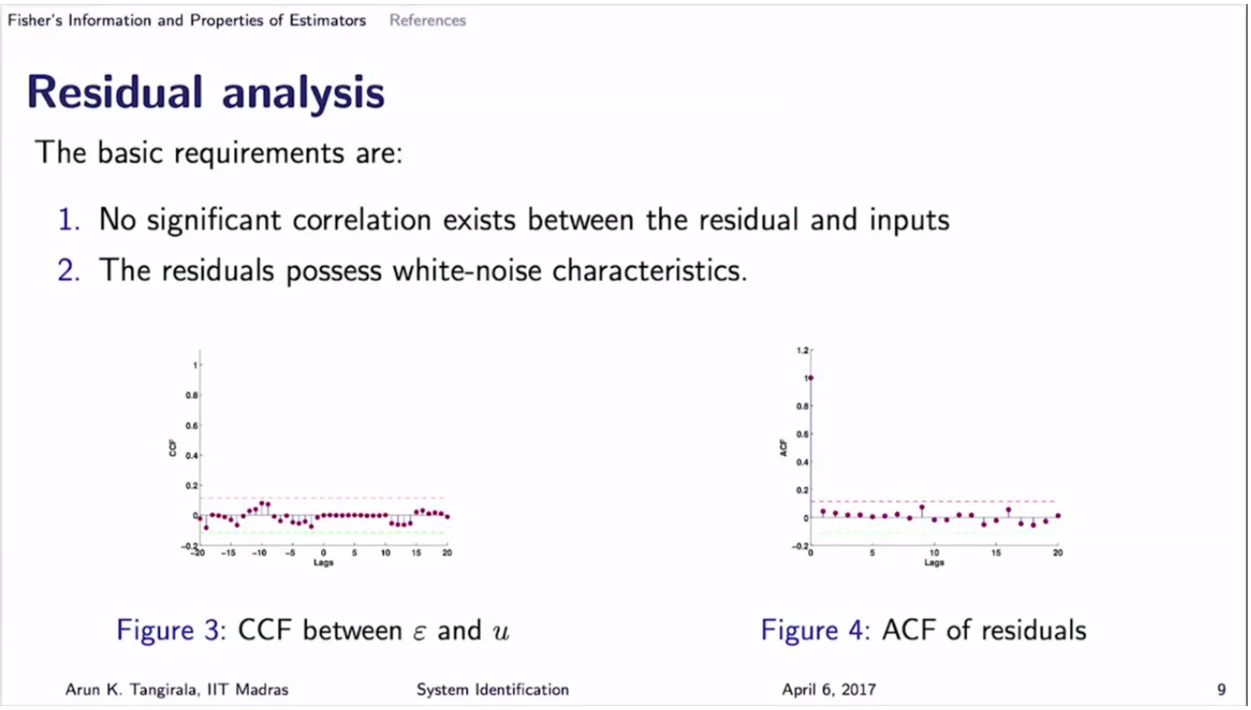
So if R square is very high, what does it mean? That means you have done a good job of fitting the data. Butthere it doesn't tell you whether you have over parameterized or done anything. It doesn't tell you anything of that sort. So many people are happy looking at R square measure and they go and have a party. But you have to be careful. You may have ended up over parameterizing. R square doesn't reflect that.

So there an adjusted R square that reflects this as you have learned in the lectures. There is an adjusted R square that penalizes for the number of parameters that you have included in the model. And that is why in the adjusted R square there is this p which is a number of parameters in the model that appears. So that as the p increases, the adjusted R square goes down. But then you have to keep calculating this adjusted R square for each p.

For this problem we have the R square value and the adjusted R square fairly high. Now what is a high value? There is no recommendation. But when you look at the number is a 90% fit, yeah. But this is not a foolproof way of ensuring that, you have predicted everything that you could. There is no foolproof, this is not a foolproof way. What is a foolproof way of ensuring that there is nothing more left to be predicted, to be captured? Residual analysis.Right? So look at the residuals.

And in system identification or even in general in fitting, the first thing we want to ensure is that, there is nothing, there is no correlation left between the residual and the regressors. If there is, then that means I will-- I have not included sufficient regressors. And what is the second requirement? What is the second requirement? There are two tests that we perform on the residual, right? Auto correlation to see if residuals are white. And it's important to do that because as you must have learned, you've learnt in the lectures and also as we'll restate that, least squares estimates are consistent and efficient only when the errors are white and they are uncorrelated with the regressors.

(Refer Slide Time: 29:45)



Fisher's Information and Properties of Estimators References

Residual analysis

The basic requirements are:

1. No significant correlation exists between the residual and inputs
2. The residuals possess white-noise characteristics.

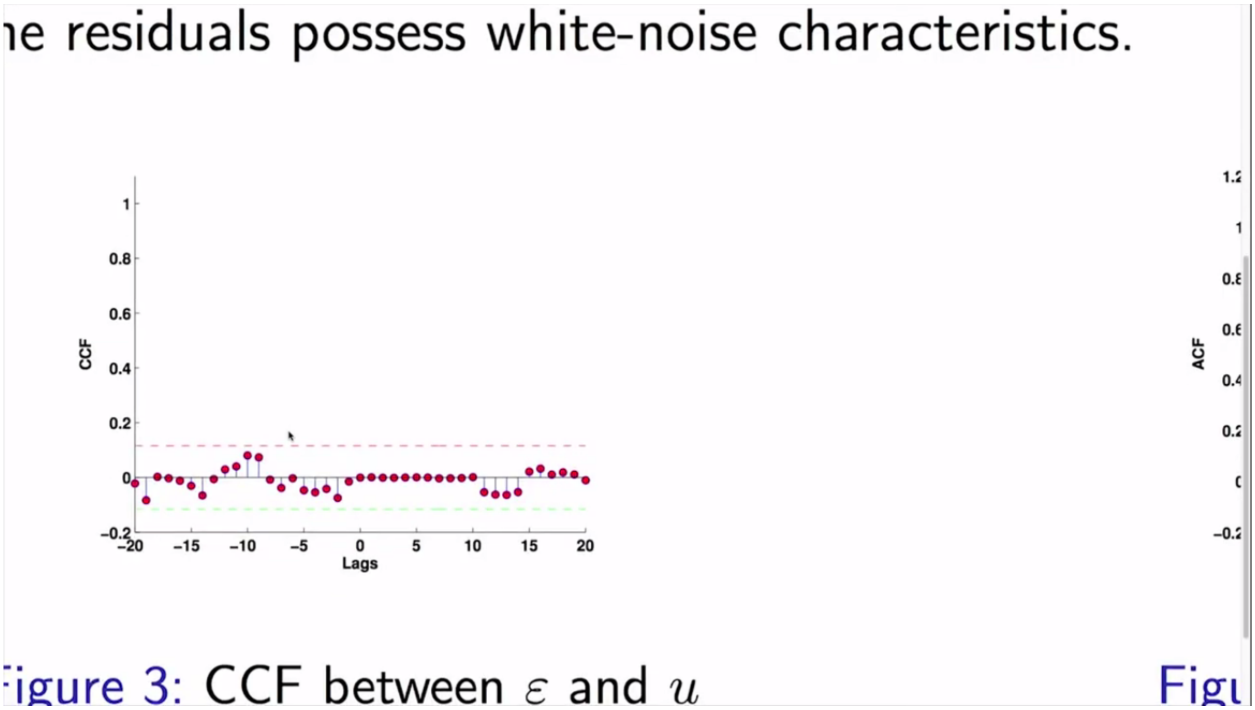
Figure 3: CCF between ε and u

Figure 4: ACF of residuals

Arun K. Tangirala, IIT Madras System Identification April 6, 2017 9

So we have to do both tests.So here is a residual, I don't know how well you can see. Let me zoom in for you. Here are those two plots.

(Refer Slide Time: 29:52)



ıe residuals possess white-noise characteristics.

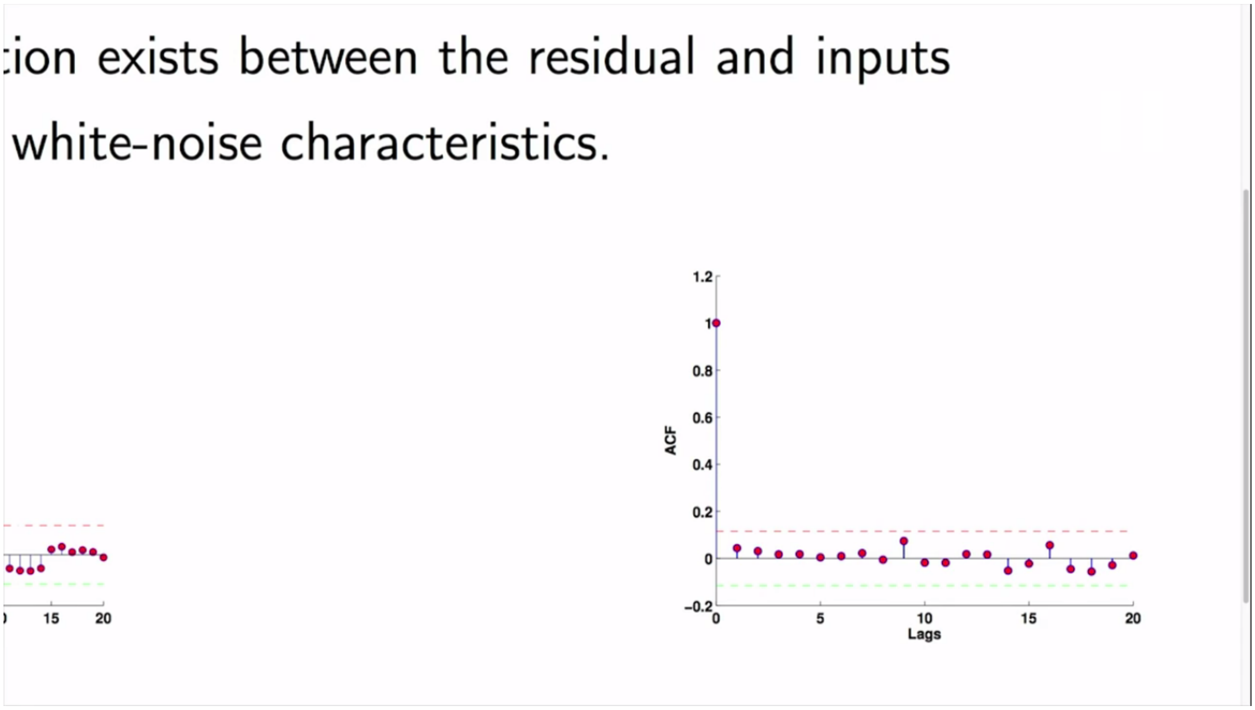
igure 3: CCF between ε and u

Figu

On the left hand side is the cross correlation plot between the input and residuals. Why am I plotting cross correlation, what does it cross correlation compute? It is computing sigma epsilon u at lag l. That means it's looking at covariance between epsilon k and u k minus l for various lags. Of course, I'm plotting cross correlation, I've written on the board cross covariance.

What should you expect to see, if you've done a good job. Here are the significance bands, I've already explain, what inaudible significance bands are. The cross correlation estimates are within that band. Basically allowing us to declare that there is no significant correlation between the residual and the regressors. In this case what are the regressors? In this example? Past inputs, that's all. Okay? So the regressors will keep changing depending on the problem, this is what I'm trying to tell you, time and again. So it says basically that there is no past effects of input present in the residuals. Which is good, that means I don't have to improve my regressor matrix.

(Refer Slide Time: 31:18)



:ion exists between the residual and inputs

white-noise characteristics.

And here is the ACF of the residuals again confirming that the residuals can be thought of as white. If the residuals-- if in my data generating process had used a colored noise, in our example we have used white-noise. If I had used colored noise, then I would have seen some correlation here. Then I have to sit and build a noise model. At the moment that need is not there. That needs will typically arise in assignments and exams. But in the example right now, we don't need to do that. Alright, but the anyway FIR model is a very basic model. Later on, we know that, we will have to build more sophisticated models, alright. So we are convinced now that this model has not under fit, right? Now we need to compute the over fitting part and for that we need to compute the errors, essentially we have to look at bias, variance and so on.