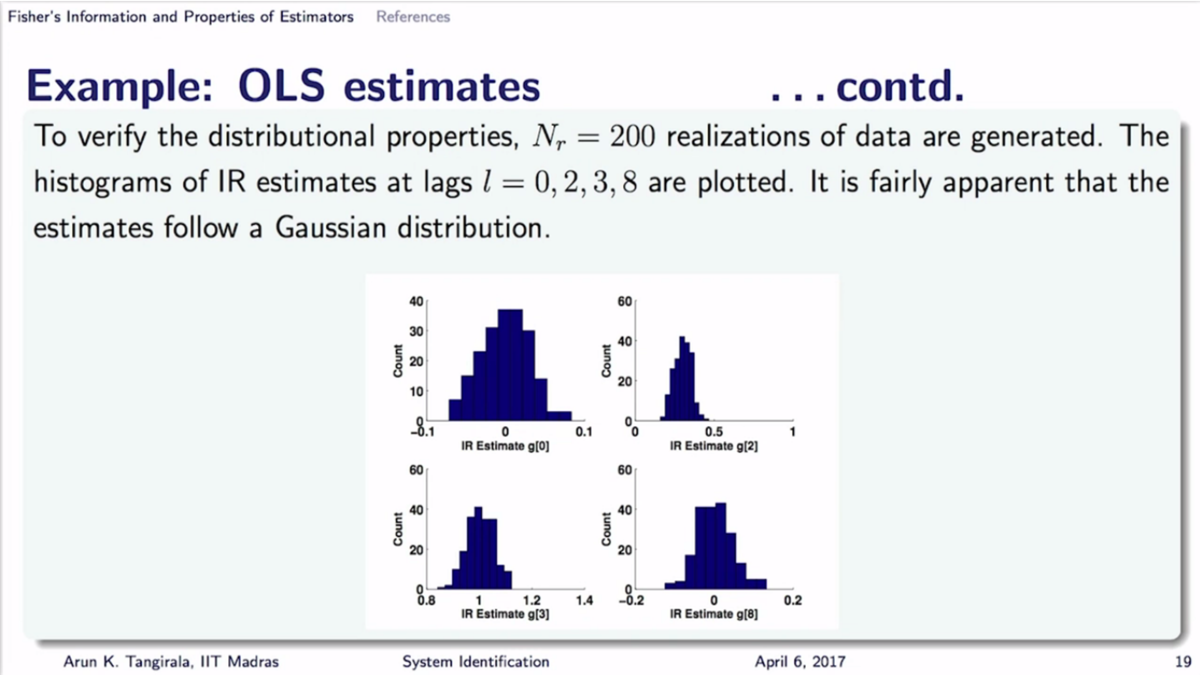
CH5230: System Identification

Fisher's information and properties of estimators

Part 11

So what we'll do today is. We will just recap some of the least squares [0:27 inaudible] material. When it comes to estimating the models G and H. We have already gone through an example in the last class.

(Refer Slide Time: 00:36)



Where we looked at estimation of FIR model and we talked about the distribution of the estimates, the efficiency of the estimates. We also went through examples telling us when the least squares estimates can be expected to be efficient and consistent and so on. So the key things just to recap. Least squares estimators are consistent and efficient only when the residuals are white. Okay. And uncorrelated even if. So consistency and efficiency should not be confused. You can have a consistent estimate but not an efficient estimate. So the scenario in which such a case occurs is when the residuals are not white. But at the same time they are not correlated with the regressors. So let's say you have an FIR model estimation case, the data generating process is also FIR but the noise corrupting the y or the v k or g k, whatever you want to call it is colored. Right?

So typically when we assume this data generating process, as an example, [02:08 inaudible] Suppose you have [02:19 inaudible].This is the data generating process. Model you have assumed is this. Assume that you have rightly gotten the number of impulse response coefficients in your model. What is a difference per say between the model and the process? The nature of the noise that you have assumed, that is, there is a mismatch in the noise model. In this case you will get consistent estimates. That is not an issue. Weekly consistent estimates in sense, in the probabilistic sense the coefficients will converge to the true values.

But efficiency is not guaranteed. Because least squares estimators are efficient only when whatever you are leaving behind. You may be assuming e k. So why do we assume this to be [03:39 inaudible]. We have to make some assumption on the noise we have made it to be white. And then you write the predictor and set up your least squares problem and estimate your b. Right? In this case, whatever the leftovers after you estimate are not going to be white. Clearly, because I know that the data generating process is not generating the data that way.

In the DGP these are the ones that are generating the data. So that the residuals are going to be colored. That's not an issue. As long as they are uncorrelated with the regressors consistency is guaranteed. v consistency is guaranteed. But efficiency is not guaranteed. You understand the difference between consistency and efficiency. For consistency, that means for your estimates to convert to the true values. It's quite important that the residual should be uncorrelated with that regressors. Otherwise you will have a bias as we have discussed. Right?

In fact I went to a simple example. Although I didn't show you the Monte Carlo, averages from the Monte Carlo realizations. It was there, we're just one step away. Even from the point estimates it was kind of fairly obvious that there is a bias. Here is a scenario where thatresiduals are not going to be correlated with the regressors but they are going to be colored. Right?In which case you are not guaranteed, efficient estimates. Again how do we know whether all of this horse residual analysis?

So once they fit this model I will not know this. This is hidden from me this is what I have with me. How do I know if I made the right assumption on the noise model? Go through a residual analysis figure out do things.Right?what are the two things that we perform in residual analysis. Cross-correlation with the regressors and autocorrelation. Both will tell me whether these two conditions are satisfied. That's it. So the cross-correlation will tell you, okay consistent is not so much of an issue but the autocorrelationwill tell you efficiency is an issue.

So how do you address that efficiency? One way is to go back and refine your noise model. You should not touch the deterministic part because that has kind of been taken care of. And there, then you are getting in to parametric noise models and so on. As it stands it's an FIR model that you're assuming, even otherwise it's an FIR but it's an FIR model with some non-white kind of noise model. Then you will have to turn to parametric model estimation and you will have to jointly estimate the noise and the deterministic model. Right?

One way is to do a loose fitting. That means, you're happy with the deterministic part. You only take the residuals. You look at the ACF or PACF,it'll tell you whether the underlying noise model is moving average or AR and so on. Fit a noise model to the residuals. So you have separately an FIR model, the deterministic part and separately the stochastic part. But that is [06:55 inaudible] jointly optimal. You will have to estimate both of them simultaneously. Which is what we do in parametric model estimation where we will estimate G and H.

But it doesn't take too much effort to guess that you would be then solving a nonlinear least squares problem. Okay. It's not going to be easy. Because it's like,see this, you can think of this as an ARX model. Right? An FIR model is also, whatever I have written here on the board. You can also think of this as an ARX model. Am I, right? What is the A polynomial here?One. Correct. Right? So this is an ARX model in fact. Now if I sit down to model the noise then. And suppose I decide, I figure out from the ACF that, yes that is a moving average company. Then I would be fitting [07:57 inaudible] model. Right? And we know ARX model leads to nonlinear [08:04 inaudible] and as a result I end up running into the nonlinear least squares problem. That is one way.

The other way that you can do is, looking at weighted least squares. And I'll briefly talk about that shortly. But before I move on to weighted least squaresand nonlinear least squares,that is that kind of constitutes or crux of 60% of today's lecture or you cansay 50% and then we look at MLE. So that we'll close and then well tomorrow will take up the estimation of all models. We have impulse response and parametric model families where we'll learn a unified method called prediction error.

So quickly before I move on toweighted least squaresthere is one point that I still want to drive home which is that you should not confuse a model structure and the estimation algorithm. For a given model structure there may be many different estimation algorithms. You should be clear in your mind where the shortcoming is. Whether it is a model structure that has a shortcoming? Or whether it is an estimation algorithm? So here, suppose I want to estimate. I mean this is the model that I have. We have just not discussed that we will not obtain efficient estimates.

Right? But I can keep this model structure intact and still obtain efficient estimates by choosing a different estimation algorithm. Yes. My model is falling short. In what respects? In the noise modeling. So I don't have to necessarily go and fix the noise model to get an efficient estimate of the deterministic part. I can still retain this model structure but use a different estimationalgorithm. That is a point that you want to remember. Sometimes you may have the right model structure but a wrong estimationalgorithm.

And that may get you [10:12 inaudible] estimates. A classic example is a moving average model fitting. Suppose I give you only noise time-series modeling. And I ask you to fit a-- you've discovered from ACF. You've gone through one such problem in your assignment. Right?Where I gave you time-series data. Hopefully you looked at ACF and PACF and figured out what models are appropriate and so on. Suppose in one such situation you figure out that moving average model is appropriate for the given series.

No input is given.Pure time-series model.Now you choose, let us say to fit a moving average model or order two because ACF is giving you such a guess. Now you feel use method of moments to fit the parameters to estimate the parameters of the moving average model. You will actually run into highly inefficient estimates. There the-- Maybe the process is [11:13 inaudible]. You have rightly guessed the order. Everything's fine. The model structure is not falling shorter in any aspect.

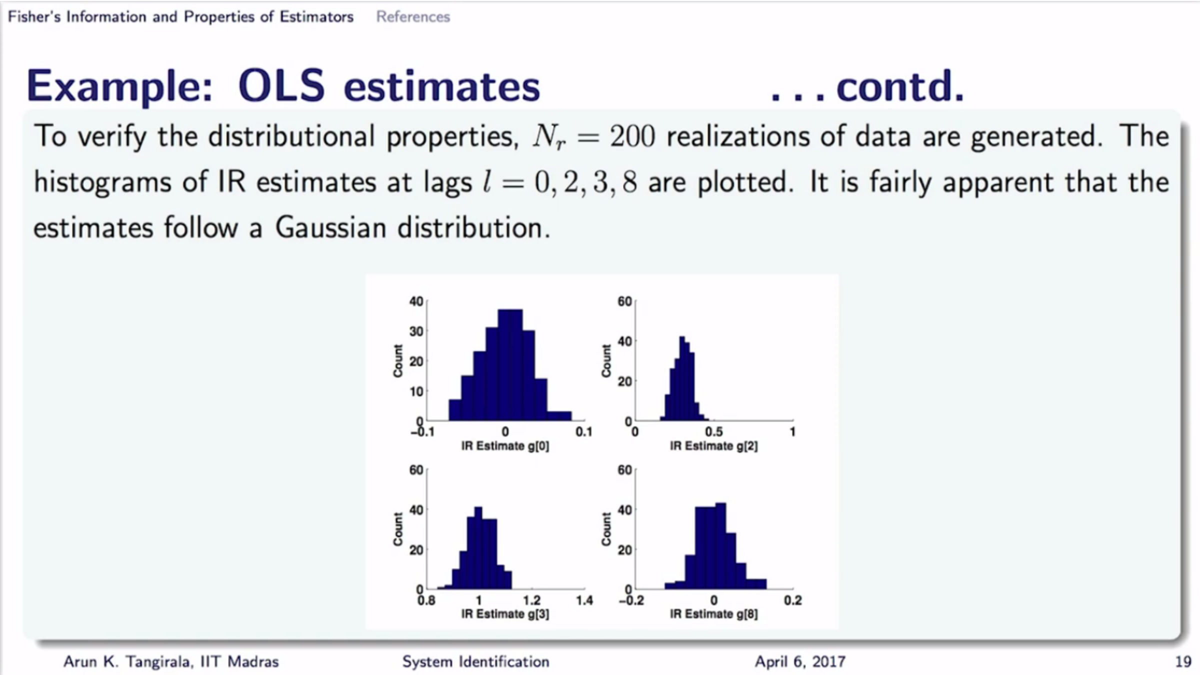
However the method of estimation, you have not chosen appropriately. Do you follow. On the other hand if you least squares or even MLE. And you'll definitely have to use a nonlinear least squares because you're going to run into a nonlinear predictor with moving average models. So, if you use nonlinear least squares or MLE you'll get efficient estimates. So you should learn to distinguish between the model structure or the model itself, and the estimation algorithm. And that is why we have been studying them separately.

Understand the model itself. Understand estimationalgorithm by itself. And now you have to decide whether you should marry them. For a given model structure which is the best.[12:09 inaudible] estimating autoregressive models. Many times we do that. Then you can use the method of moments which gives rise to Yule-Walker equations. In fact for autoregressive models you can either use method of moments, you can use least squares or MLE all of them give asymptotically same estimates. In the sense same efficient.

So you see there the model structure combined with estimation algorithmmakes a difference. For always you have to learn to remember this. When that identification or in any modeling exercise the user should be well-versed with the nature of the model and the properties of an estimator. If you're not, then you may not only end up giving a wrong model but a very poor model. Poor model in the sense, rightmodel structure but poor estimates.

So keep that in mind. And that is why we are going through a different estimation algorithms, their properties particularly. So with least squares, what are the things that now you ought to remember.

(Refer Slide Time 13:24)



One is that you want to make sure that the residuals are white and uncorrelated with their regressors. Now I have been using the FIR model example. All right.But when we move into the parametric model family, you will also have the past values of output coming in. Right? Then your regressors are going to be constructed from past outputs and past inputs. Suppose I had an ARX model, regular generic ARX model then I would have had past outputs as well. So your regressors there are past outputs and past inputs of appropriate lags.

When we talk of correlation of cross-correlation between the residuals and the regressors you do not have to, in such cases evaluate the cross-correlation between residuals and passed outputs and residuals and passed inputs separately. Why? It sufficient to only look at cross-correlation between residuals and inputs. You understand the point. Here I have anFIR model, so the natural thing to do is after I fit the model I look at the cross-covariance function or cross-correlation function. That part is easy to see because you want to. See if the residuals have correlated with the regressors. Clear?

But [14:54 inaudible] Right? Let's say I am going to fit this model here. We tried using a green chalk on a green board but I don't mention. Okay. Use [15:08 inaudible] model plus b 1, let say u k minus 1 plus e k. So here, again what I meant earlier is, you do not have to look at two different cross-covariances. This is just because you have both outputs and inputs as your regressors. You don't have to look at these. Ideally you supposed to look at the cross-covariance between residuals and regressors. Right? The regressors here are passed outputs and passed inputs. Of course, here I have only one of them, one passed output and one passed input. You don't have to look at both. It is sufficient to still examine only this. Why?

Like in a residue plot, I mean, when you use a residue command to perform your residual analysis, it brings up two plots, only two plots. One is thecross-covariance of the residuals with the inputs and is the other one, autocorrelation. Why doesn't it bring up this one?Anyway.Yeah. It's generated by the input itself. So the effects of inputs are contained in y. What am I interested when I'm doing this residual analysis? I just want to see if the model is deficient in some respects. That is what I'm looking at.

I'm not trying to figure out in residual analysis whether I should include y k minus 2 or uk minus 2 or you know, some other lagged output and input. I'm not interested in that. And usually you should not rely on residual analysis for that. To figure out which term you've missed out. Only under some very specific conditions. You can figure out which term is missing and usually those conditions are met by the academician in a classroom session. Okay. In all other real life situations you will not run into the scenario where you can figure out which term is missing out in you model.

So the goal of residual analysis is not to determine which term is missing. Rather the objective is to figure out whether you have some, whether G is falling short of something or H is falling short of something, whether under fits have occurred. In the plant and noise models.That's what I want to know. Yeah. There are some other, one or two other indications that residual analysis can give you we'll not talk about it right now. But what you should definitely remember is, not to use a residual analysis to figure out which terms are missing.

You understand?When I say residual analysis,particular thecross-covariance. And the general order of procedure is, first you make sure that your deterministic model is doing well. Right? That it has been satisfied. And then you turn to the noise model. Right?And then you figured out whether you have left out anything even there you may have chosen maybe an MA(1) or an ARMA(1,1) and so on.
And the fact that the residuals are auto correlated will only tell you that, "Yes, your model is falling short. Still you will not be able to tell you whether you should increase the AR order or the MA order.

All right. So order determination is not whatresidual analysis is meant for. That traditionally has been done through the use of trial and error combined with information criteria methods. But now today what we have is, a host of other methods based on Spars optimization but will not get into that in this course where you can automatically determine the order. State space identification which we will learn next week. We'll give you a good handle on order determination but again under some ideal conditions.