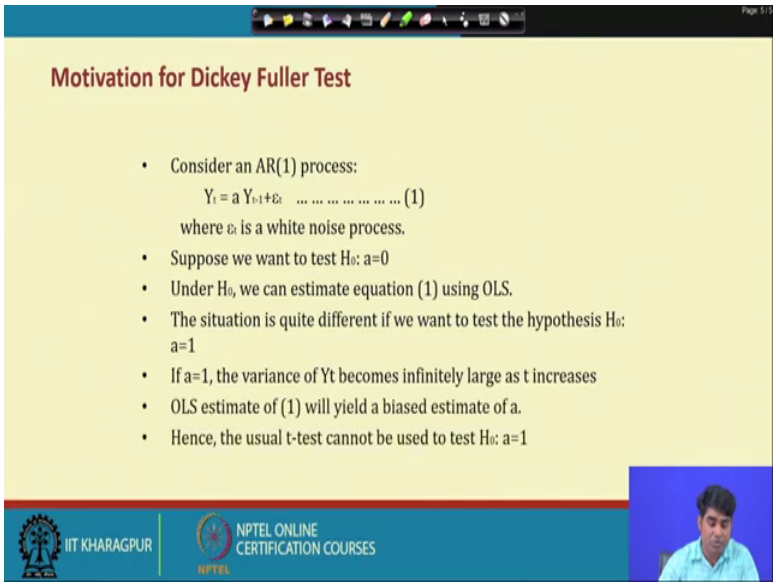
Engineering Econometrics
Prof. Rudra P. Pradhan
Vinod Gupta School of Management
Indian Institute of Technology, Kharagpur

Lecture – 52
Time Series Modelling- VAR modelling y

Hello everybody, this is Rudra Pradhan here, welcome to Engineering Econometrics and we are in the process of Time Series Modelling in that too the structure of you know VAR modelling. In the last lecture, we have discussed something called as you know unit root, then co-integration and then the structure of you know VAR, we have specifically highlighted, what is the structure of you know unit root and the kind of you know co-integration. Once again, I am just connecting to the unit root and then move to co-integration and finally the VAR setup.

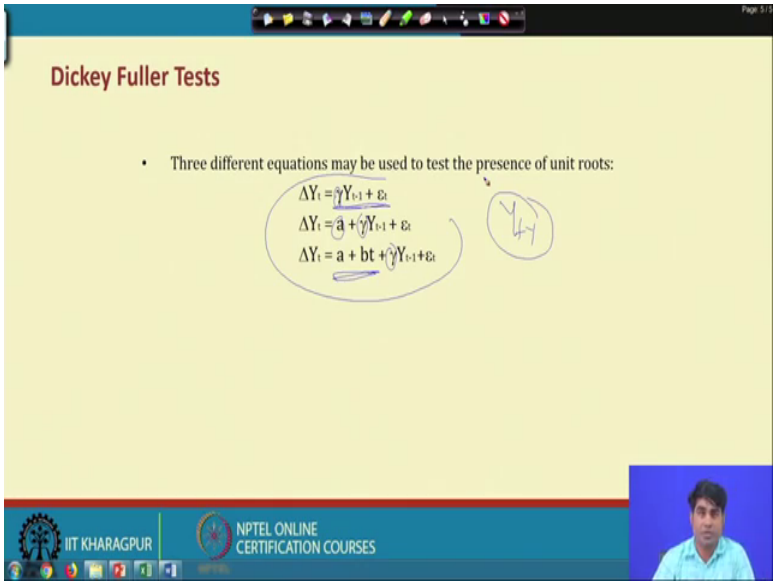
(Refer Slide Time: 01:05)



So, the structure of you know unit root start with you know autoregressive scheme like this, so, where Y t is connected with Y t minus 1 and we are supposed to check actually whether a is statistically significant or not. So, this can be tested with you know without any constant and trend, we can test with a constant and we can test with constant and trend so; that means, every time there are 3 different ways, we can test the unit root. So, the Dickey Fuller test is like this and we have to check whether you know, the particular

coefficient is statistically significant and you can go ahead with this simple OLS technique and then derive the conclusions ok.

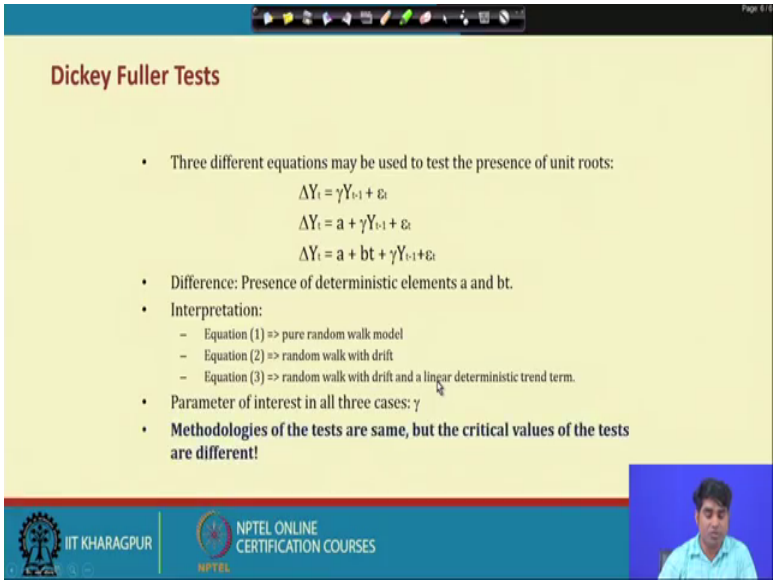
(Refer Slide Time: 01:58)



Then if it is actually statistically significant then we will start with rejecting the null hypothesis that the variable is actually non stationary and ultimately the variable will be stationary. If not then you will regress the model with you know first different equations and as a result. So, we have a here 3 forms, the first one is with no constant and you know trend so; that means, we have simple start with the variable so here, there is a constant and here there is a constant and trend.

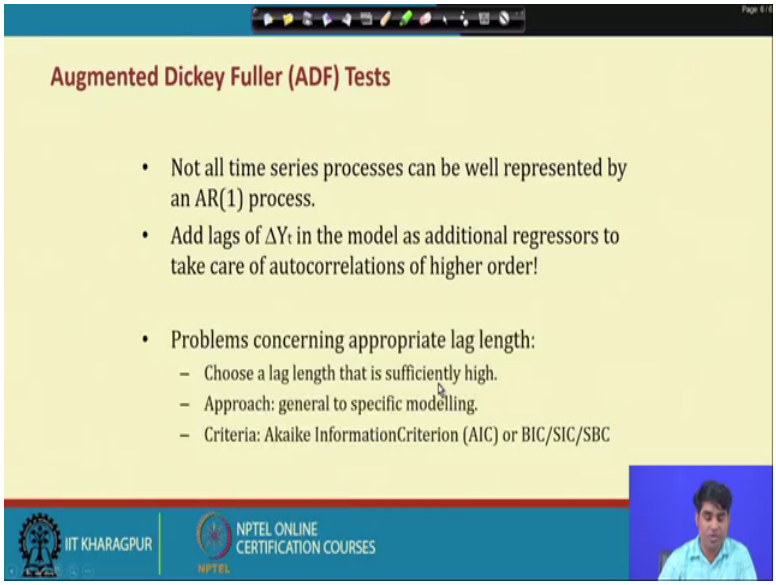
So, whatever maybe the forms ultimately, the end result is you know the statistical significance of these parameters, the coefficient of you know Y t minus 1. And that brings the nature of the stationarity and the order of integration that is by this simple Dickey Fuller test.

(Refer Slide Time: 03:06)

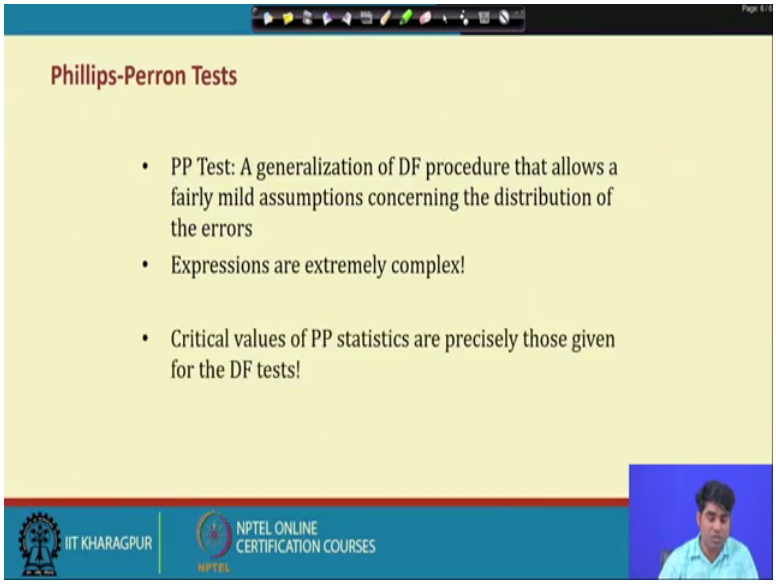


So, again I am not going in details and then you like to check actually, whether the variable is stationarity or not. And if it is stationary, at what levels and again, in what forms; is it with you know pure random, walk without drift and without trend and again random walk with drift, that is what the constant and then random walk with drift and you know trend that is you know dealing. So, the ADF is just extension to the DF by addressing or you know adjustment with you know the lag length. So, likewise we have a another test called as you know Phillips Perron test.

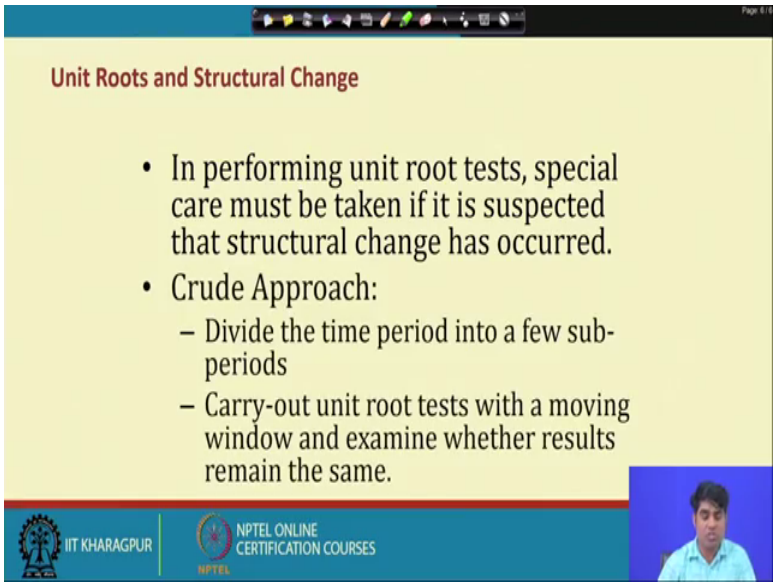
(Refer Slide Time: 03:34)



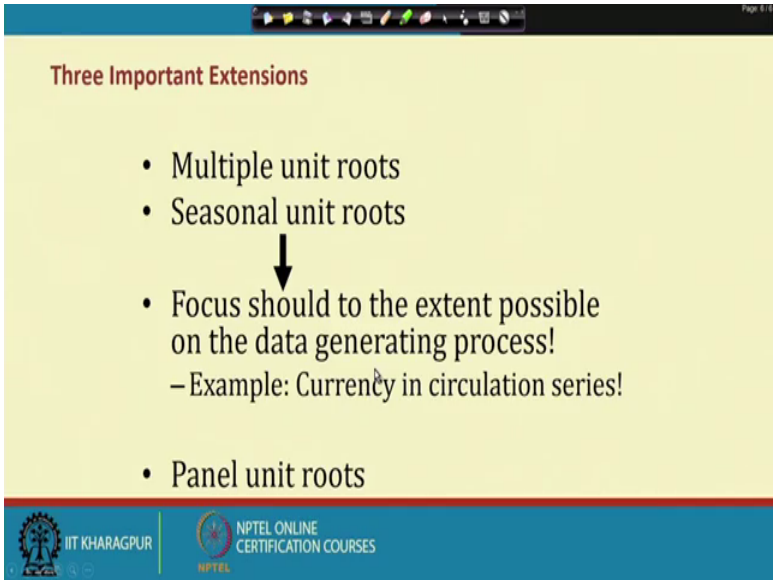
(Refer Slide Time: 03:44)



(Refer Slide Time: 03:48)

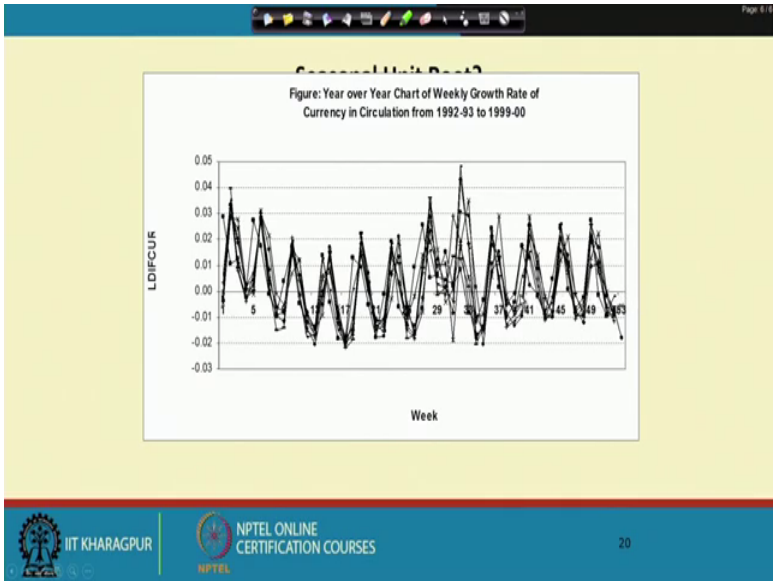


(Refer Slide Time: 04:03)



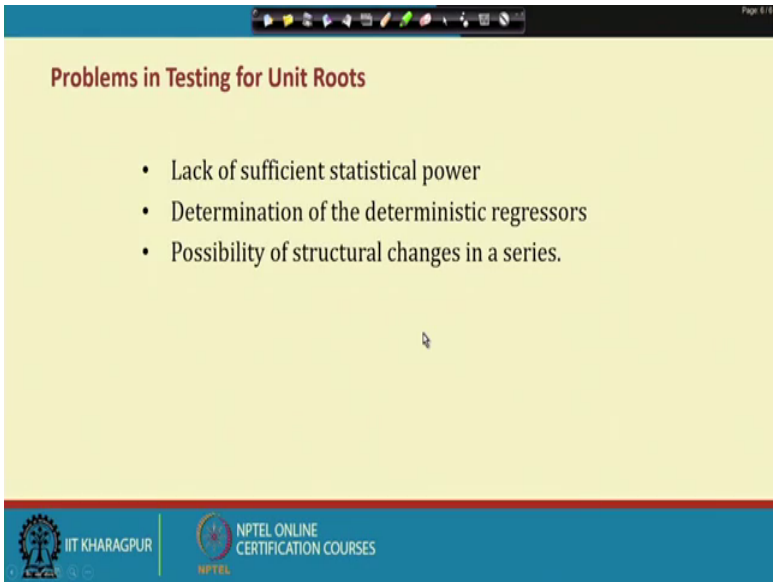
And some of the structural change may happens so in that case, the best test statistics which you can apply is called as a unreliable test. And unreliable test will give you some kind of you know structure to carry out whether, the variable is reaching stationarity or not.

(Refer Slide Time: 04:10)

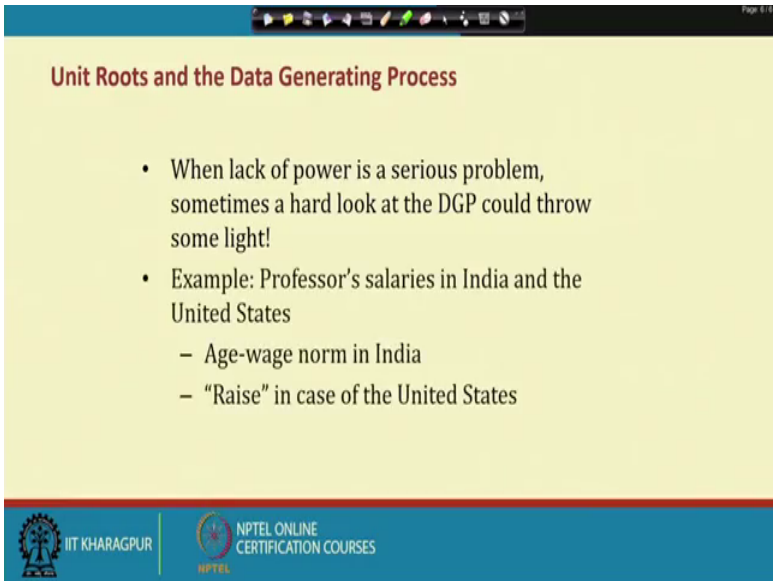


So, now, what we will do here? So, you know we will just check the behaviour of the variables and bring whether the variable is reaching stationarity and at what forms right.

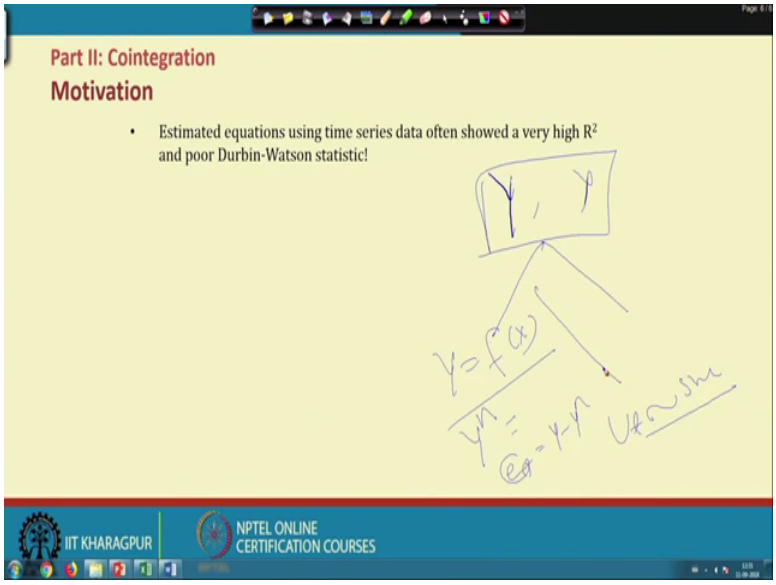
(Refer Slide Time: 04:19)



(Refer Slide Time: 04:22)



(Refer Slide Time: 04:26)



So, after that we will move into the concept called as you know co-integrations, let me bring into the co-integration concept, because the concept of unit root, we have already highlighted in the last lecture and the co-integration starts with you know simple, regressing Y upon X provided these system starts with you not integrating Y and X. So for instance, let us say there are 2 variables Y and X. And, we like to know whether Y and X are correlated each other, there are various you know methods are there starting with you know simple to complex.

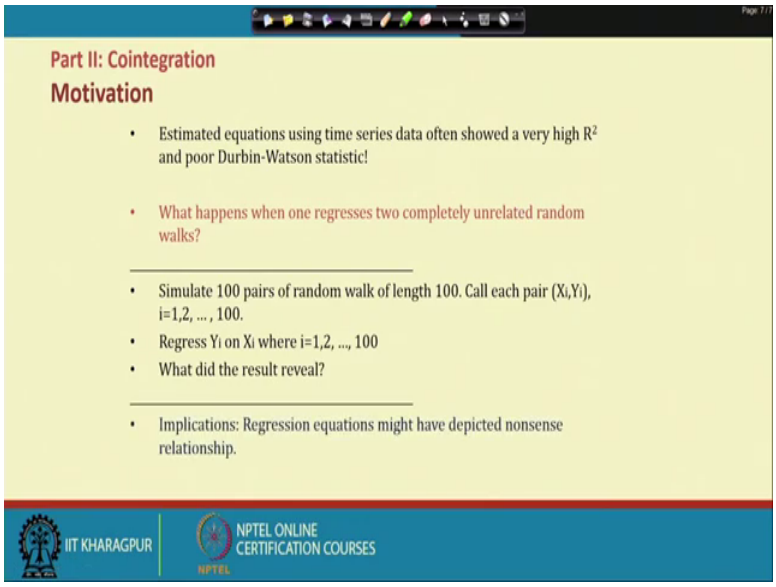
So, we can actually use the a you know simplest form of the model called as you know Engle Granger's approach E G approach or you can apply simply called as you know Johansen Juselius approach. In the Engle Granger approach, which is actually simplest way to know the things it has the 2 step process, that is why it is called as a Engle Granger's 2 step co-integration test.

The first step is to check whether, you know Y and X are you know related to each other. So, we start with you know regressing, bring linear regressions in that too through LS technique and then check whether beta coefficient is significant and get the estimated equation. And finally get error terms, which is difference between Y and Y head and we need to check whether, the particular error term is actually stationarity or non stationarity, that is the step 2 process.

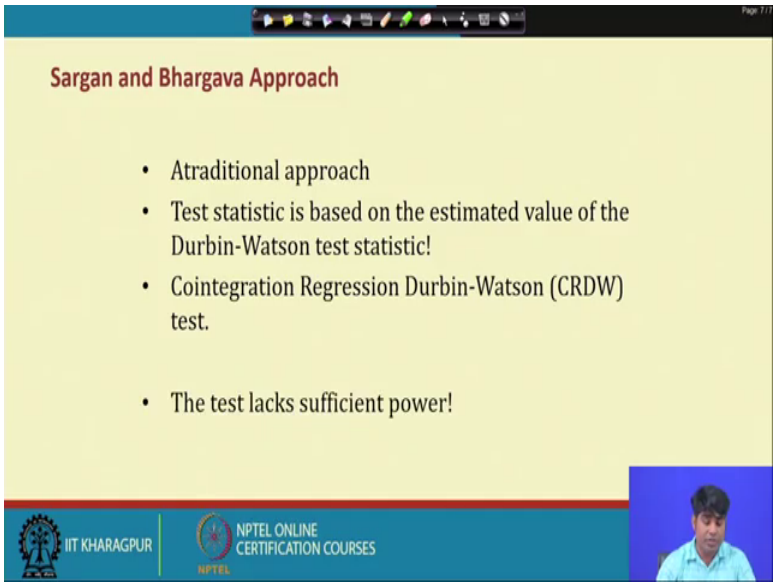
In the step 2, step 1 the coefficient should be statistically significant and in the step 2, the error variable should be stationarity in nature, if that is the case and we will have it then,

we declare that you know Y X are you know co-integrated to each other, that is the declaration we will have and that is the need through, which you can actually a proceed for the vector autoregressive schemes.

(Refer Slide Time: 06:29)



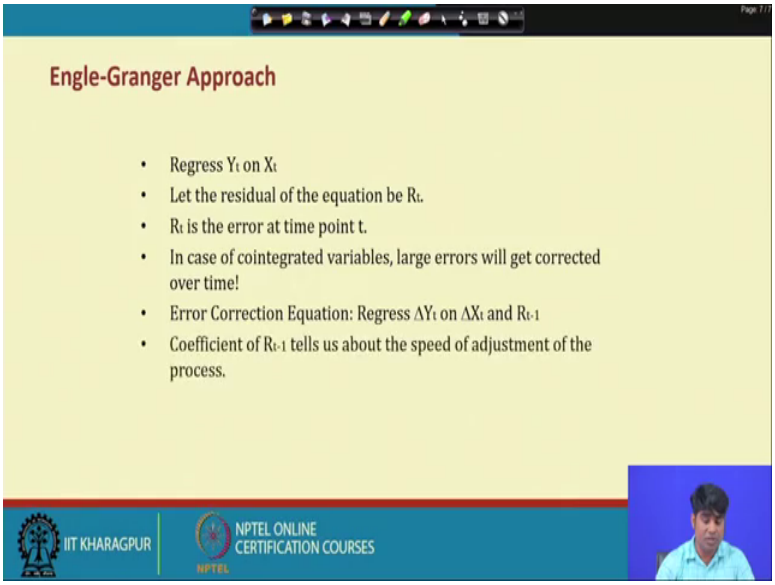
(Refer Slide Time: 06:38)



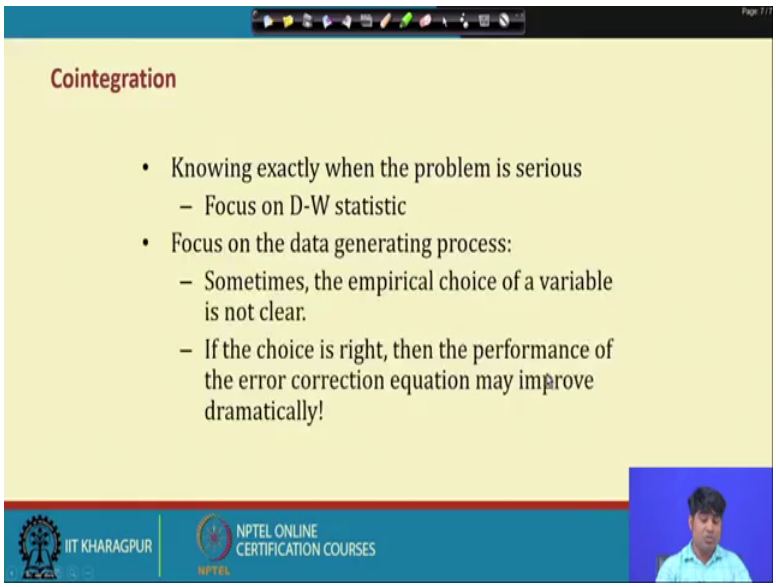
So, what we will do actually? So, this is what we know called traditional approach, then you know even if we can use co-integration regression by Durbin Watson, that is technically called as a CRDW and then we w ill go for by you know, simple Engle Granger approach, what I have already highlighted ok. So, this is step by step process,

simply regress Y upon X then get the residuals and check residuals you know stationarity and that itself will give you the indication about the you know entry to VAR setup.

(Refer Slide Time: 06:55)



(Refer Slide Time: 07:19)



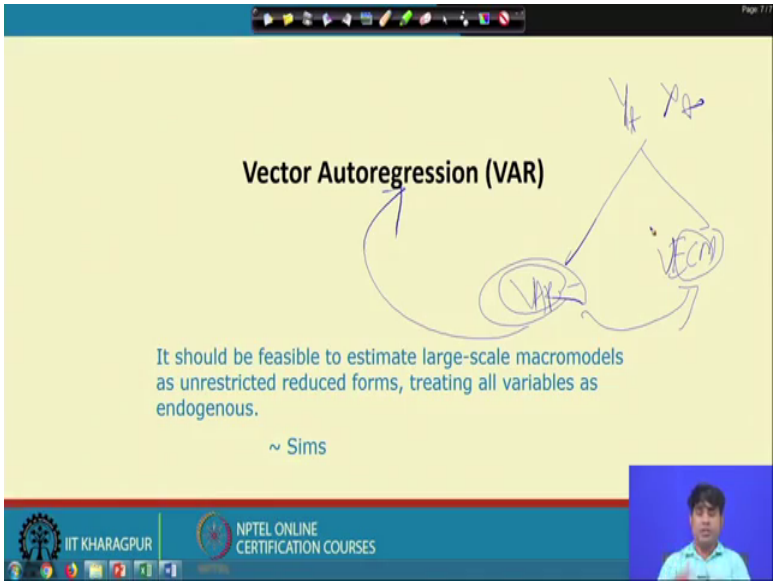
And a technically, we have also you know, different you know test structure called as you know Johansen Juselius, it depends upon you know it is again, you know it is a kind of you know vector format and like here, when we are you know using Engle Granger's approach or you know CRDW. So, where you know, it is a kind of you know simple

form, we just regress Y upon X and then derive the error term and then check the stationarity of the error term.

But in the Johansen Juselius approach, we are suppose to you are supposed to bring actually the kind of you know vector forms, where we have 2 different test statistic through each you declare the you know co-integration, that is the you know structure called as you know trace statistics and maximum eigenvalue statistics. And, on the basis of you know trace statistics and eigenvalue statistics, we can declare, whether the variables are you know co-integrated or not. In the Johansen Juselius approach, so if you look for the kind of you know co-integration, then with respect to 2 variables so there is a feasibility of you know, 2 co-integration or 1 co-integration or you know say no co-integration.

So technically, whether it is a 1 co-integration or 2 co-integration, there is a co-integration and if not there is no co-integration. So now, the VAR you know structures specifically you know needs, the kind of you know the kind of you know inference about the co-integration.

(Refer Slide Time: 09:14)



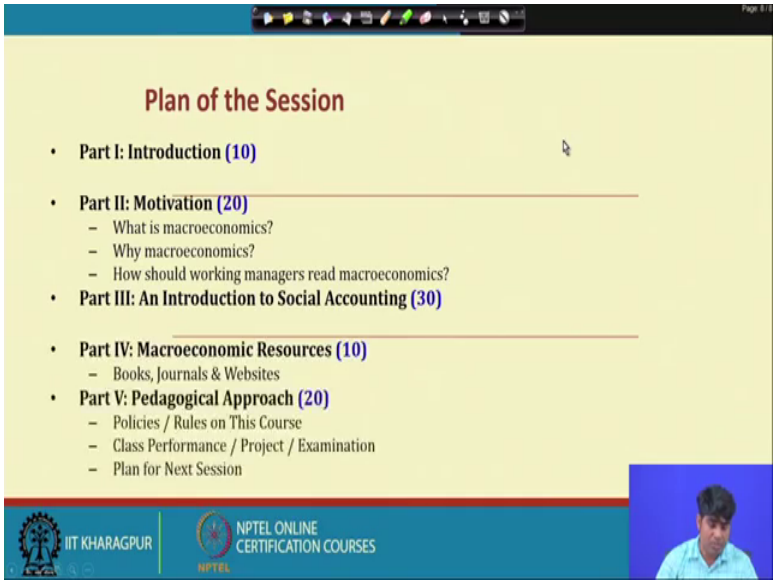
So technically, in the VAR setup, if you go to the VAR, the VAR setup is like this. So, we like to know what is the, you know linkage between Y and X? Whether you know unidirectional, bidirectional or simply neutrality and we can you know test this

hypothesis through simple VAR structure or if not then the VECM structure. So that itself, you know differentiate with the help of you know co-integration.

So; that means, technically the issue of you know, the issue you know VAR is like this. Technically it will be like this so for instance, let us say Y and X and we call it Y t and X t and then we start with you know VAR setup and then there is called as you know VECM setup, vector autoregressive which is actually the core of this problem. And, then the other form is called as you know VECM, that is the first VAR and then ECM is the error correction mechanism so; that means, there is a error correction term which we will derive because the usually the error terms U t is entered as you know error correction terms by bringing the form called as you know U t minus 1.

And then as usual the VAR models and then we introduce the error correction term and then again we go for the estimation. So here, you need to check you know the coefficient of the error correction terms as well as the coefficient of you know variables, the coefficient of the error term will give you the long run single on and the coefficient of the you know variables will give you certain signal. So, ECM involvement means so there is a indication, that you know co-integration, but that co-integration may not be actually in the long runs. So now, in order to know whether there is long run kind of you know inference. So, the ECM term will be error correction term will be integrated with the VAR setup, then the model can be turned into vector error correction model and then finally, you have to estimate and test the kind of you know structure.

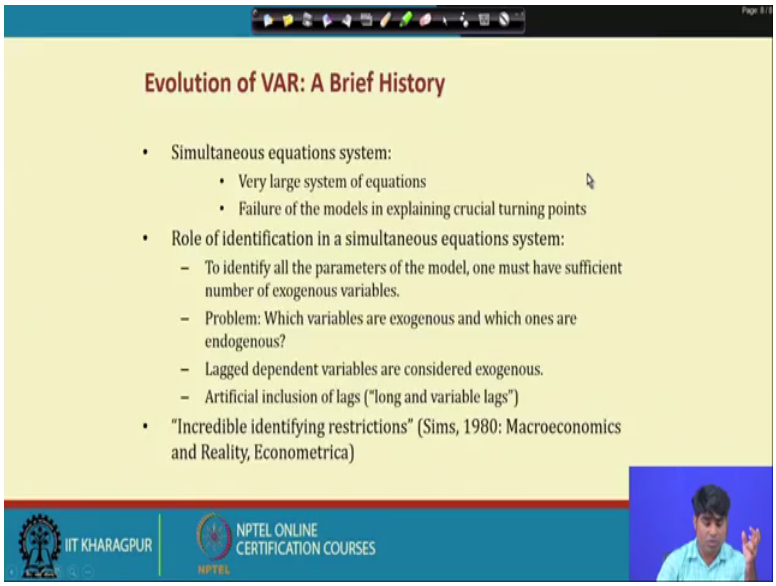
(Refer Slide Time: 11:37)



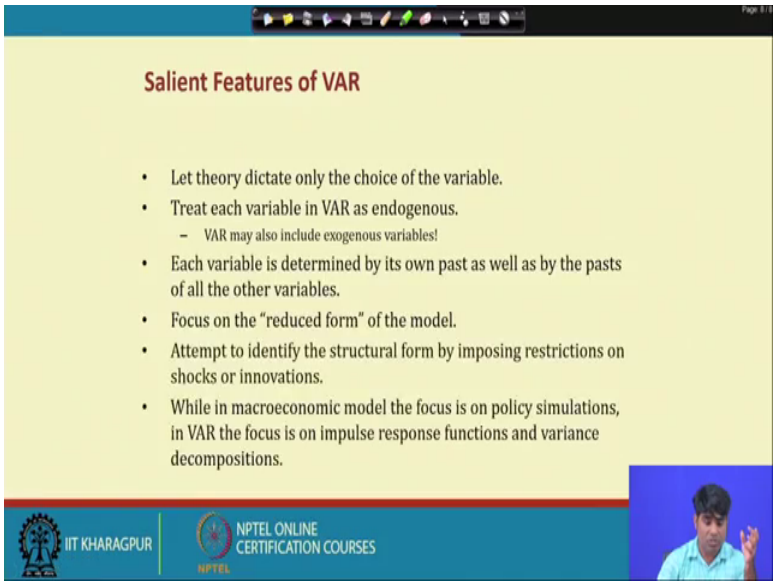
So, according that is what is the actually structure. So accordingly, we can see what are the things we are suppose to discuss? So, the usually most of the you know financial data or you know time series data in engineering time series data. Basically actually, you can say that you know financial engineering, most of the variables is having actually times spread and; that means, informations are available with respect to different point of time.

And, this kind of you know model is very easily useful, there like you know arch model and GARCH model and autoregressive and moving average and then. So, according that is what is the actually structure. So, accordingly we can see what are the things, we are suppose to discuss? We will we like to you know check the kind of you know inference, there is lots of you know history behind this you know VAR setup.

(Refer Slide Time: 12:09)

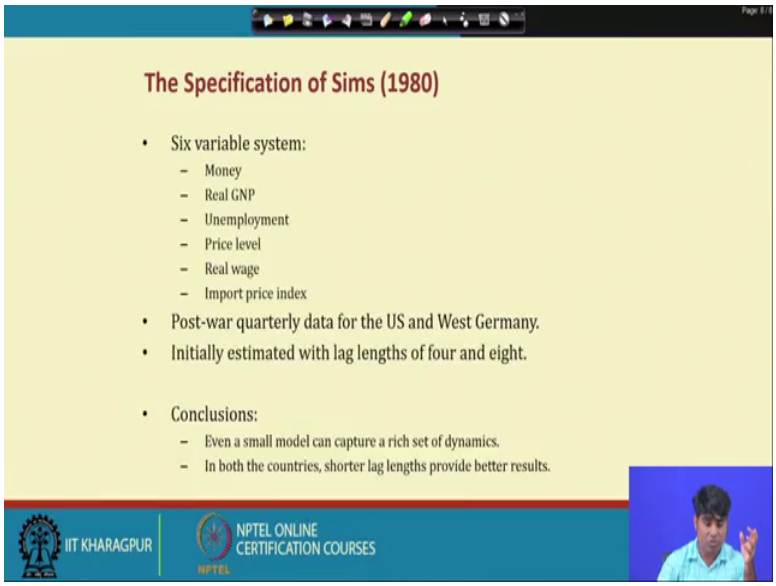


(Refer Slide Time: 12:18)

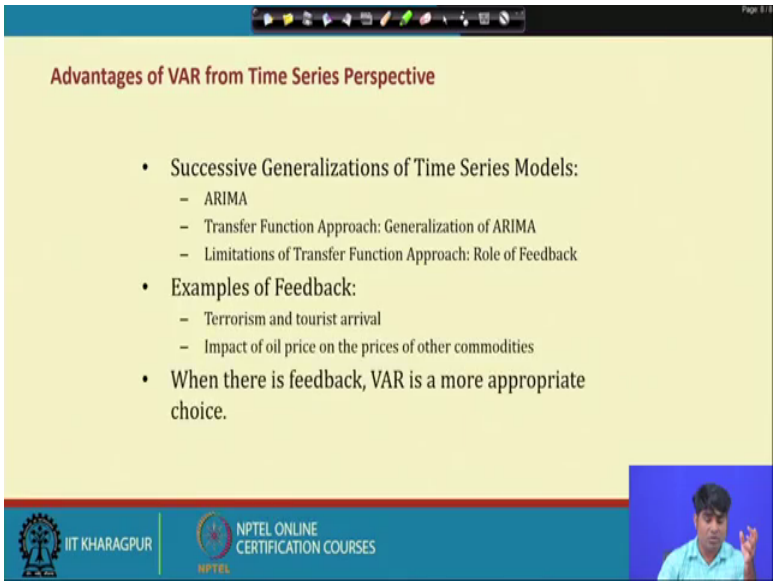


So, we can actually simply skip this.

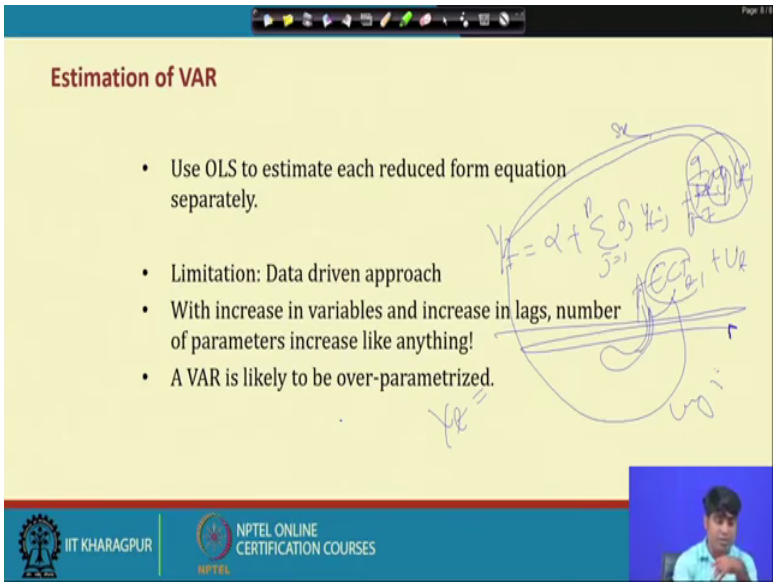
(Refer Slide Time: 12:22)



(Refer Slide Time: 12:26)



(Refer Slide Time: 12:31)



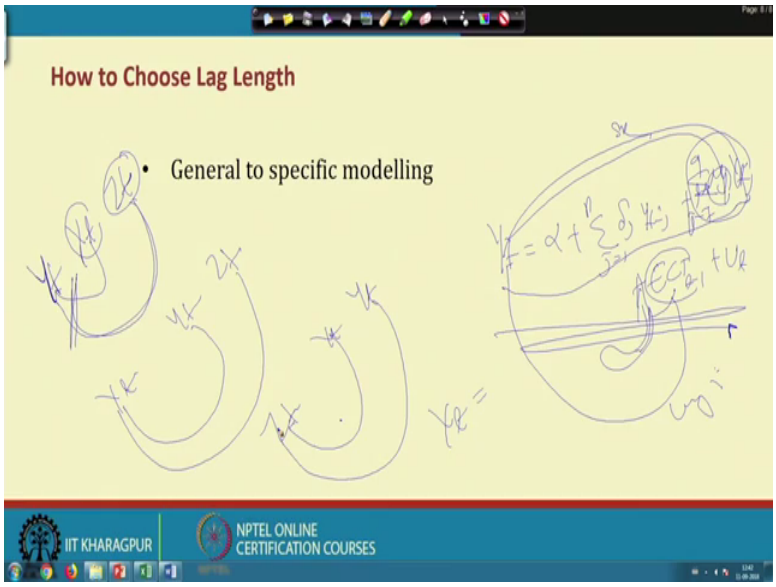
And then we can bring into the kind of you know structure, where we can have the, you know requirement ok. So, the VAR form is actually start with a simple again OLS mechanism, because the structure of VAR model will be like this, you see here Y t as a function of you know alpha plus summation delta j Y t minus j plus mu summation mu j X t minus j, j equal to 1 to p let us say, j equal to 1 to q. Let us say and plus a error correction term ECT t minus 1 that is and then error term another error term, that is what the model is all about.

And here and this so, we can put it here actually, let us say lambda the coefficient to error correction term and we need to check whether this is statistically significant and in addition to that, whether you know the coefficient of you know mu j is statistically significant. So, this brings that you know X t is influencing Y that too in the short run and this significance of the terms to Y t brings the long run impact ok. So likewise, you can develop another model for X t same way and again you have to check the coefficient of Y t here and the error correction terms then the given signal that it is the Y t which influence the X t.

So that means, technically with 2 variables, we have 2 different you know model, one with respect to Y t another with respect to X t and then check separately and come into the conclusion, that you know whether both the models rejecting the 2 null hypothesis to justify that X is causing Y and at the same time Y is causing X. If that is the case then,

we can say that you know that there is a bidirectional causality, if only one happening the other one is missing then it is called as unidirectional causality and both are missing completely then it is called as you know what we call as you know neutrality happenings. So, the thing is that you know, it is tested with you know 2 variables only.

(Refer Slide Time: 15:15)



But in the meantime, this can be extended with you know multiple variables so; that means, the structure will be let us say a 3 variables, we know we can start with like this you can start with like this in a say Y t ok. Let us say, Y t X t and Z t ok. So now, again the structure will be here, Y t starts with the Y t minus j then X t minus j and then Z you know Z t minus j. So, that will be another, variables to be added into the system and then as a result, we have a 3 different models like this. So now, you like to check X t upon Y t Z t upon Y t again, while you know, where you know we are you know targeting the first equation, that times X t minus Z coefficient Z t minus t coefficients will be these you know variables of you know choice to predict the Y

And again when we integrate with X t and Y t as the reference and Z t is the reference then, that times we targeting whether Y is influencing X and Z is influencing X. Again so, Z t is the reference variables, where we are linking X t upon you know Y t of course, Z you know lag of this Z t will be there, here lag of the X t will be there and here lag of the Y t will be also there and then check, whether X t is influencing Z t and Y t is also influencing Z t. If that is the case, then we like to check how is this kind of you know,

linkage? So; that means, we have a multiple options here, compared to you know bivariate case, in the bivariate case we have specifically 4 different situations, unidirectional Y to X, while the reverse is not true, then again unidirectional X to Y the reverse is not true then bidirectional, where both can go each other and neutrality that do not cause each other, that is what the kind of you know concept.

But in the case of you know trivariate here. So, we have actually 3 such combinations, Y X, Y Z and X Z and again we have 4 different specification under each groups. So technically so, if 4 different you know kind of you know inference in each group. So, technically we have a 12 different you know kind of you know inference inferences, which we are supposed to be checked and you know get to know, how is the kind of you know setup. Ultimately the VAR requirement or you know VAR structure is a you know just to know whether the particular variables are you know interdependent to each other and that too in a time series framework, it is slightly different to the ordinary, you know linear regression model even though we are using here you know OLS mechanism, there also we are using OLS mechanism.

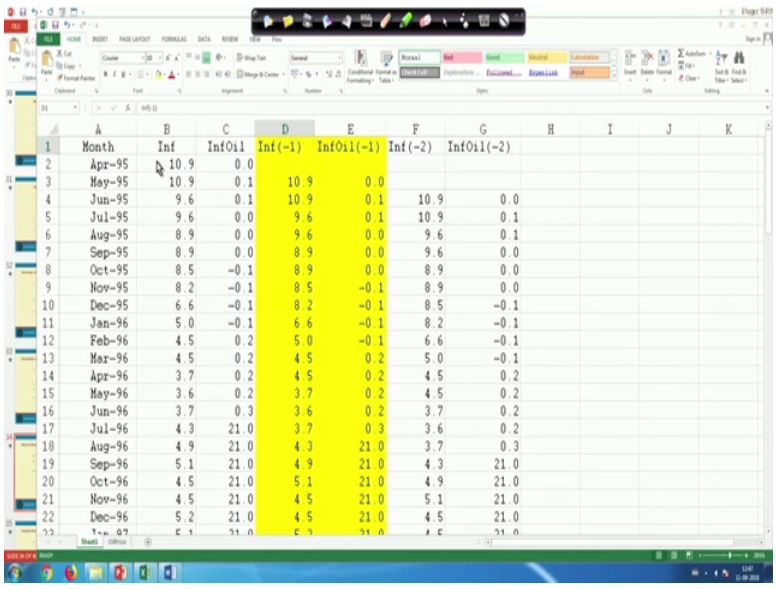
In that case, we are just integrating Y t upon X t and Z t. So, Y t cannot be another variable, independent variable there and all are you know independent, just we are you know regressing X t and Z t with Y t, that is the 3 variables, there in the process, but here the case is slightly different. Here of course, there are 3 variables, but by spread with lag. So, it can have you know n number of variables. So, Y t minus j X t minus j, then Z t minus j again, we have no clear idea, how the j will you know will vary, is it with respect to 1 or with respect 2, with respect to 3, the moment you start you know 1 1, 2, 3 lag then you know model spreads will start increasing and that brings you know, simple model to complex models and simple structure to multivariate structure.

Even though, we are starting with you know VAR still this model has lots of you know complexity and the kind of you know extension. So obviously since, you know there are lots of you know lags involve in this process by default, we need to first optimize the lag length. So, optimum lag length is the first hand choice in the VAR setup and that is how the idea is you know to see, how to choose you know lag length? That is that is the big deal. So, we have already discussed the choice of lag length, in the case of you know arima cluster, arch cluster and GARCH clusters and where we have actually specifically

highlighted that you know, there are certain test statistics this includes AIC statistics, SIC statistics, finance prediction error statistics.

These are the statistics are also equally used here to optimize the lag length and that is the first and criteria of this process, once you do this then; obviously, by default you can fix the lag length first. And after fixing the lag length then, you come to a kind of you know conclusion that, you know model is you know specified, correctly specified and then look for the kind of you know causality inference. For instance you know, the model starts with at least you know 1 lag fast, it cannot start with 0, it will start with you know 1 lag then it can start with 2 lag, it can start with 3 lag, 4 lag and so on. So that means, we have lots of flexibility here, we have lots of flexibility here, this same inference can be checked with respect to fixing lag length 1 again, you can check these inference with fixing lag length 2, again check the inference by fixing lag length 3 and so on.

(Refer Slide Time: 22:03)



	A	B	C	D	E	F	G
1	Month	Inf	Inf0il	Inf(-1)	Inf0il(-1)	Inf(-2)	Inf0il(-2)
2	Apr-95	10.9	0.0				
3	May-95	10.9	0.1	10.9	0.0		
4	Jun-95	9.6	0.1	10.9	0.1	10.9	0.0
5	Jul-95	9.6	0.0	9.6	0.1	10.9	0.1
6	Aug-95	8.9	0.0	9.6	0.0	9.6	0.1
7	Sep-95	8.9	0.0	8.9	0.0	9.6	0.0
8	Oct-95	8.5	-0.1	8.9	0.0	8.9	0.0
9	Nov-95	8.2	-0.1	8.5	-0.1	8.9	0.0
10	Dec-95	6.6	-0.1	8.2	-0.1	8.5	-0.1
11	Jan-96	5.0	-0.1	6.6	-0.1	8.2	-0.1
12	Feb-96	4.5	0.2	5.0	-0.1	6.6	-0.1
13	Mar-96	4.5	0.2	4.5	0.2	5.0	-0.1
14	Apr-96	3.7	0.2	4.5	0.2	4.5	0.2
15	May-96	3.6	0.2	3.7	0.2	4.5	0.2
16	Jun-96	3.7	0.3	3.6	0.2	3.7	0.2
17	Jul-96	4.3	21.0	3.7	0.3	3.6	0.2
18	Aug-96	4.9	21.0	4.3	21.0	3.7	0.3
19	Sep-96	5.1	21.0	4.9	21.0	4.3	21.0
20	Oct-96	4.5	21.0	5.1	21.0	4.9	21.0
21	Nov-96	4.5	21.0	4.5	21.0	5.1	21.0
22	Dec-96	5.2	21.0	4.5	21.0	4.5	21.0

If your sample size is indefinitely very high then, you can you have lots of you know flexibility to has the kind of you know models. For instance if you go to the, you know problems excel sheet which we have already highlighted earlier, you will find excel spreadsheet, where we have already used some kind of you know, analysis for the unit root. So now, in this case see you know for instance the simple VAR modelling and what we like to do? See let say Inf is the dependent variable here, then it will be regress with

the Inf t minus Inf with you know Inf t minus 1 then Oil t minus 1 provided inflation and Oil, you know oil consumption 2 are you know in the kind of you know in the basket.

So now so, the inflation can be predicted with you know fast behaviour of inflations and the oil consumptions that too vast behaviour of the oil consumption with the intention that, you know the lag of these variables will influence the current performance of you know inflation then, your if you fix you know lag length 1, optimum lag length 1, which we will actually finalize with the help of OIC and SIC statistics and then you just regress now, you know Inf with Inf minus 1 and Oil minus 1. Again, you go simple regression mechanism and regress and check the inference of course, if we use VECM then, first you actually bring, you know the error correction terms by linking actually inflation with you know Oil then get the error correction term ok.

And the lag of the error correction term, which is actually instrumental in the VECM process so; that means, technically if you bring BSM then another variable will be added into the system in the form of you know ECM and then ultimately, you are supposed to check the coefficient of you know Oil and the coefficient of you know ECM, the coefficient of Oil will bring the certain impact of Oil to inflation and ECM will bring, you long run impact to the inflations and in the similar fashion, this, this, this will be this, will be the kind of you know, this will be the kind of you know structure you can follow dependent variables and that can be linked with you know Inf t minus 1 and Oil t minus 1 and then the error correction term since, Oil t minus 1 is the dependent variable.

So, the target variable will be inflation t minus 1 whereas, the certain impact to the Oil and then the kind of you know error correction term, the coefficient error correction term, which is actually long run, long run impact to the Oil then you check the kind of you know behaviour.

Now, if you try to extend the lag length let us say you know you are not happy with you know one lag length you can go for you know 2 lag length or 3 lag length of course, we like to fix on the basis of AIC and SIC statistics still, you can actually test. So, in that case let us say if you are fixing a 2 lag length, let us say Inf is the original variable then, you have Inf t minus 1 Oil t minus 1 Inf t minus 2 Oil t minus 2 and then error correction term; that means, technically you have 5 you know 5 set of you know variables to this particular system so; that means, you just imagine you know, if you start you know

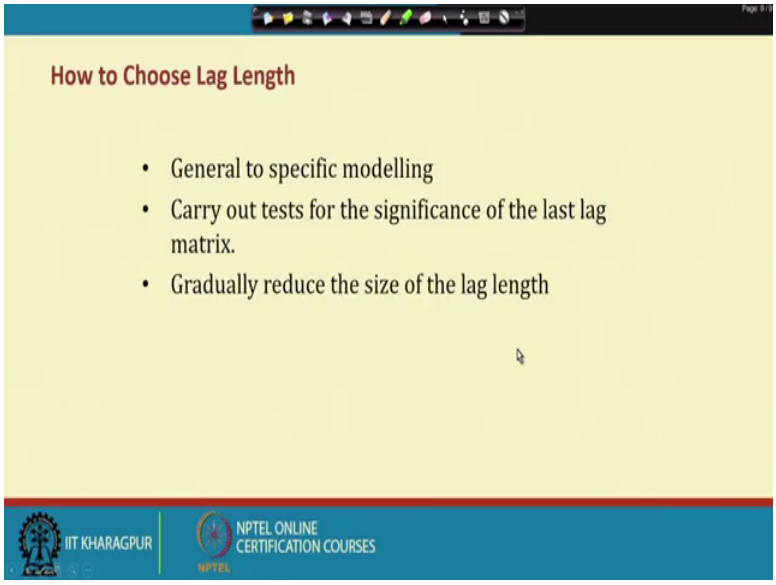
increasing lag length you know, how is the complexity of this estimation and their kind of you know model.

But it is the means that itself is the beauty of the VAR system and you know vector autoregressive system and by the way these are the requirements through, which actually you have to predict the kind of you know some of the engineering variables. Because, the behaviour of a particular engineering variables in a current form or the requirement of the future form exclusively, depends upon it is you know, first behaviour that is what is actually represented by the lag structure, which we are actually observing here. And, after knowing the particular structure then you can use any software to check, but theoretically it should first convince that you know yes the lag variable impact is there in the current you know performance of this variable then you can actually start.

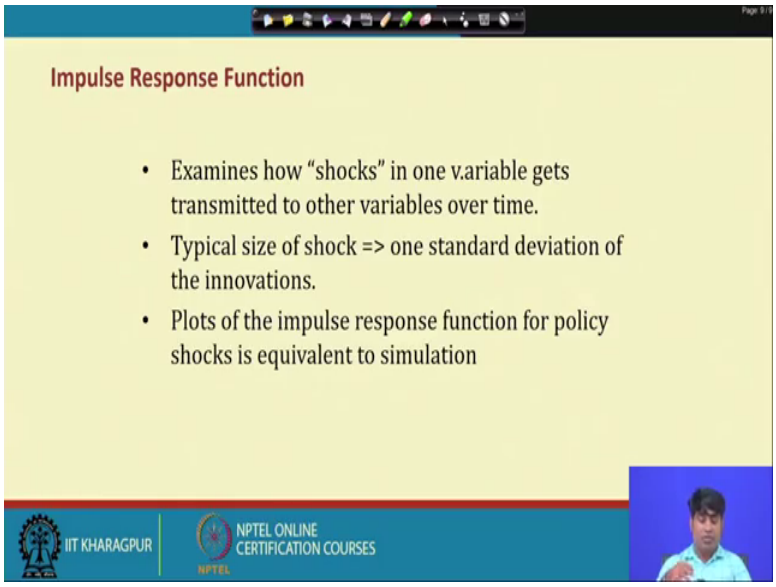
So, every time the, if we starts with you know theoretical understanding because, it is at the means end of the day it is all about econometrics. So, theoretical hint and logical hint is actually mandatory to start this process otherwise, it is a simple you know mathematics mathematical operation only which we can do easily, but the thing is that you know, since it is a kind of you know application oriented and looking for a solution to a engineering problem.

So, you should actually in the first end, convince that the theoretically there is such kind of you know linkage and you know evidence and then through this technique and data you empirically verify and being the you know, reality what is exactly happening on the basis of theory and that too what data reads about it? What model speaks about it? That is the kind of, you know structures.

(Refer Slide Time: 27:46)



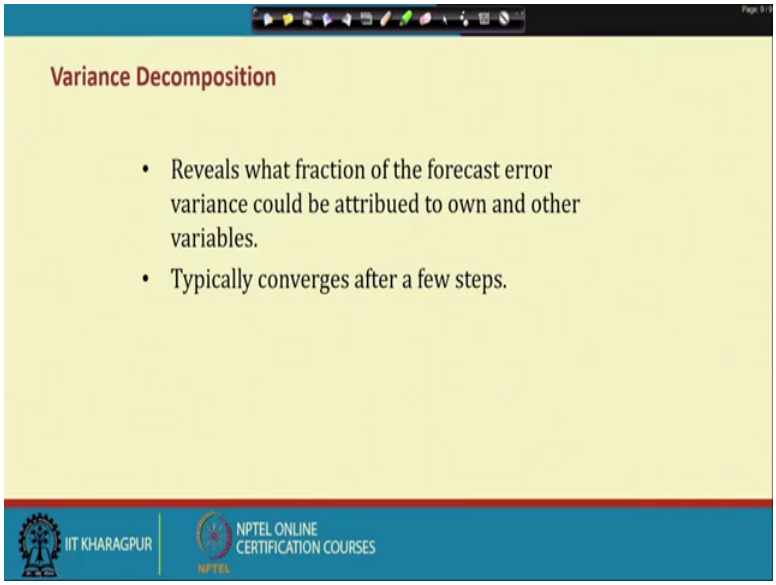
(Refer Slide Time: 27:50)



So; that means technically, I like to highlight that you know VAR is a kind of you know structure, which is very interesting to carry out with some kind of you know inference and in response to VAR model and you know vector regression modelling, which we can quantify and then we can test in a kind of you know quantitative framework by checking the coefficient of these variables and the coefficient of the error correction term. And by the way, we sometimes use actually the kind of you know chi square test to check the validity of the models ok. But in response to you know the estimation process and the check process that means I have the testing.

So, we have 2 different form, for which you can you know justify, the reliability of the VAR model or VECM model that is called as you know impulse response function, that is graphical inspection about this, you know model estimation and you know variance, you know decomposition schemes. So, through you know impulse response you like to check, how shock in one variables, gets transmitted to other variable over the time.

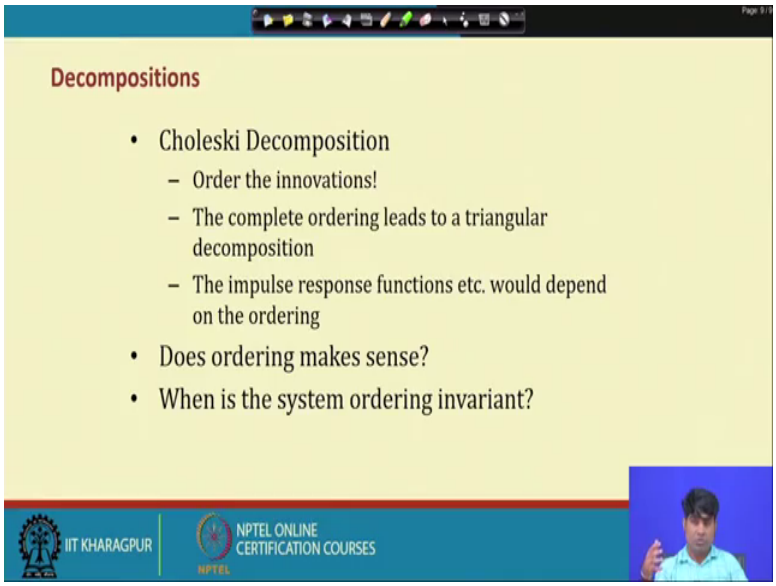
(Refer Slide Time: 29:26)



That is actually through impulse functions, we can derive and then usually the type you know typical size of this shock is actually, should be you know greater than to one standard deviation of the innovations, which is generated. And lots of these impulse response function for the policy shocks and that too it is equivalent to some kind of you know simulations right. So, the variance decomposition scheme, it reflects that you know what fraction of the forecast error variance could be attributed to the own and the other variables;

That means, we are linking actually let us say, Y t with Y t minus 1 and then the kind of you know X t minus 1 or Z t minus 1; that means so, what is the own influence that variable, you know if you are starting with Y t Y t minus 1s; that means, the variable own behaviour is also reflecting the current form. And, then the other variables lag will also influence that, you know the variance factor and the forecasting structures will be derived through variance decomposition schemes.

(Refer Slide Time: 30:12)



Typically converges after a few steps and if this is the case, then we can actually get to you know get to know that, you know the reliability of the model and some kind of you know robustness, we are you know bringing. So, you now we are just ordering the innovations, complete the you know ordering leads to a you know triangular decomposition. And then finally, the impulse response functions would depend on this particular you know ordering.

So, these are all actually different kind of you know setups through which, you know justify the importance of the VAR modelling and their reliability structure and ultimately end of the day, we like to justify that you know the VAR model is a reliable model through which actually, we can forecast certain engineering, you know problems and that too with the availability of you know some time series informations. Until unless, you have a time series data you can bring into the lag forms and then we cannot use actually VAR modelling, even if all the time series modelling starting with the autoregressive, ARIMA, ARCH GARCH and the kind of you know VAR, every times your first hand information should be a time series structure. If that is not the case then, you cannot use this models for any kind of you know engineering prediction and forecasting and then you cannot come with a kind of you know decision making situation.

So, the first hand requirement is that you must have a clear cut, you know understanding about the problems and must have variables, you know clarity, how the present you

know past form can influence the current form and the future requirement. And, then must have you know availability of you know time series data and that too in a long span and that is what the strength of this types of you know models. If your time span or the variable information is not so good enough then, this models that means, all these models cannot be very handy to predict certain things as per the particular requirement. With this we will stop here.

Thank you very much, have a nice day.