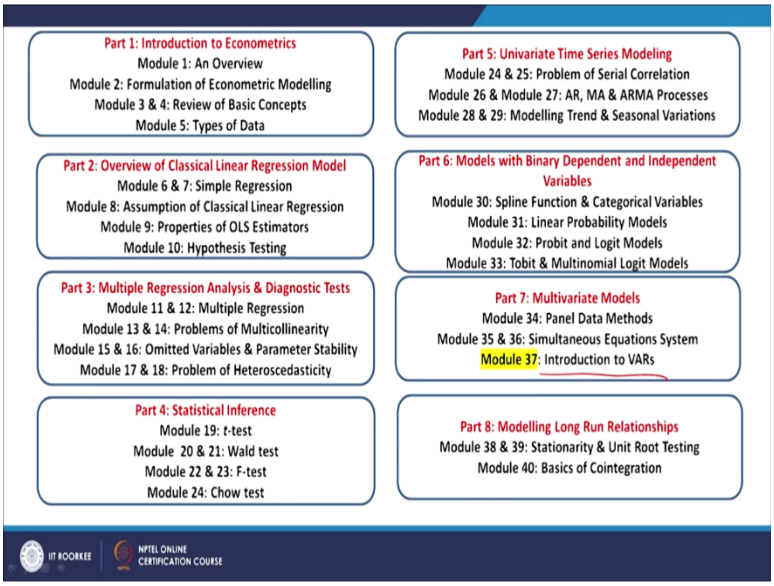
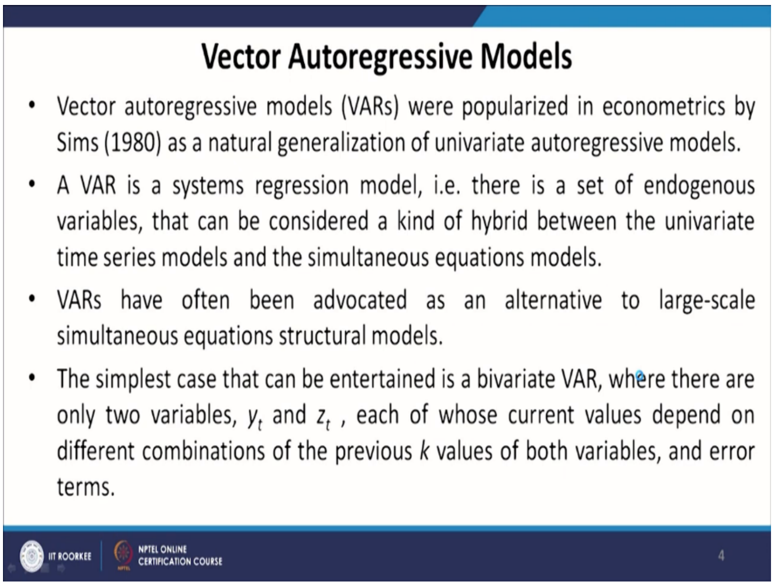
Econometric Modelling
Professor. Sujata Kar
Department of Management Studies
Indian Institute of Technology, Roorkee
Lecture No. 37
Introduction to VARs

(Refer Slide Time: 0:33)



This is Module 37 of the course on Econometric Modelling. Module 37 is a part of multivariate models, and this is sort of the last part. So, here in this module, we are going to discuss the introduction to VAR. VAR stands for vector autoregressive models.

(Refer Slide Time: 4:01)

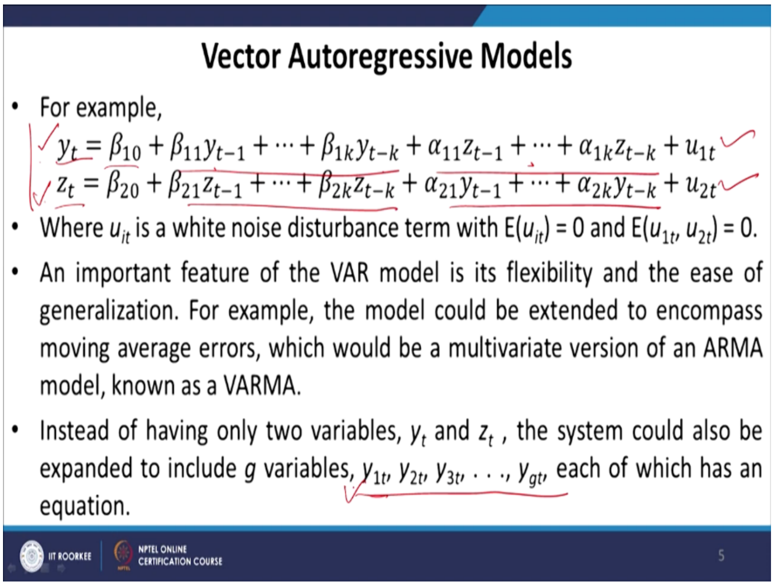


So, vector autoregressive models were popularized in econometrics by Sims in 1980, as a natural generalization of univariate autoregressive models. We have so far studied univariate autoregressive models like the autoregressive models AR models: AR(1), AR(2), and AR(P), and also we have considered the MA model moving average models.

Now, this is an extension of those autoregressive models, univariate autoregressive models, but of course, this is a vector autoregressive models are not univariate in its true sense, it is multivariate. So this is an extension of univariate series to multivariate structures. A VAR is a system regression model that is there is a set of endogenous variables that can be considered a kind of hybrid between the univariate time series models that we have studied earlier and the simultaneous equations models.

VARs have often been advocated as an alternative to large scale simultaneous equations structural models, the simplest case that can be entertained is a bivariate VAR, that is, there are only two variables where the two variables are y_t and z_t each of whose current values depend on different combinations of the previous k values of both variables and error terms. So, how do we write the VAR models?

(Refer Slide Time: 2:16)



So, this is an example where we are writing (refer slide time: 2:19- 3:08) and then we have two error terms you u_{1t} is the error term associated with the first equation and u_{2t} is the error term associated with the second equation.

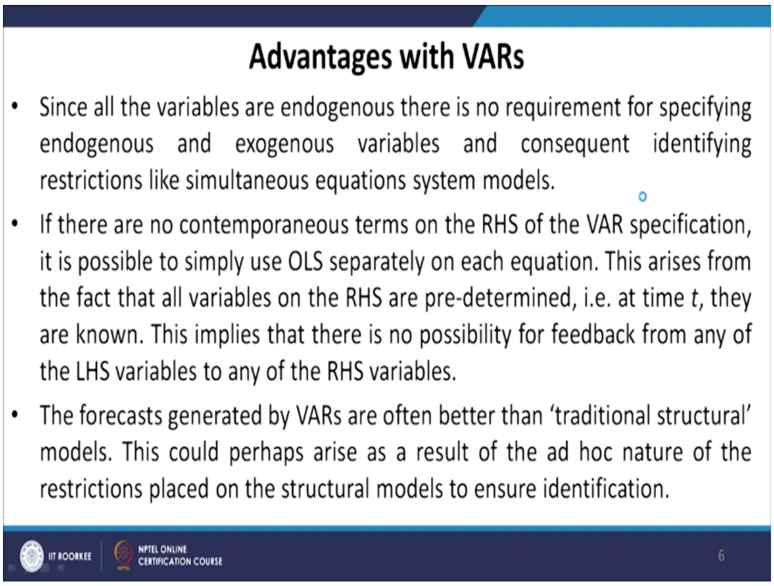
And as you can see that they are specific to our time period. So, at time period t we are considering k lagged values of both the variables and trying to explain the value of y_t or trying to figure out whether these lag values explained y_t or not. And in a similar fashion at time period t we are examining whether the lag values of both z and y explain the current value of z or not.

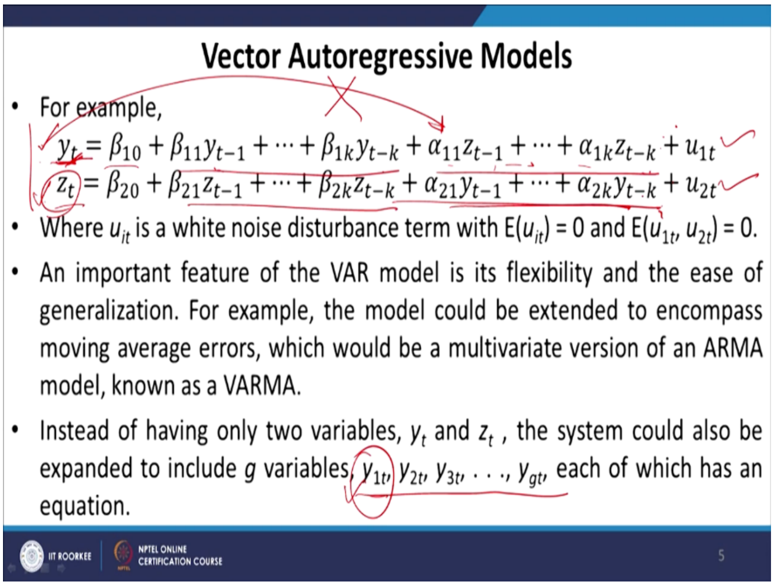
Now, this $u_{it}s$ where i equals 1 and 2 is a white noise disturbance term, and we assume it to have an expected value equals to 0, and there is no covariance or correlation between these two error terms that is the expected value of u_{1t} and u_{2t} also equals to 0. So, basically, the errors are independent of each other.

Now, an important feature of the VAR model is its flexibility, and ease of generalization. For example, the model could be extended to encompass moving average errors as well, which would be a multivariate version of the ARMA model known as VARMA. Instead of having only two variables y_t and z_t the system could also be expanded to include g variables g can be any number. So, we begin with an example of $2g$ can be 3, 4, 5, 6, 10 anything, such that, we have y_{1t}, y_{2t}, y_{3t}, ...$upto\ y_{gt}$. And similarly, we can also have several equations or several variables.

So, instead of having variables y_t and z_t now, we are having y_{1t}, y_{2t}, y_{3t}, we can also call them like y_t, z_t, w_t, X_t, and so on. So, each of which has an equation. In a VAR system, each one of these variables will have an equation, which will have lagged values of all the variables included therein, and we need to find out whether the lag values of the variables explain the current value of a particular variable or not.

(Refer Slide Time: 5:29)





So, since all the variables are endogenous, there is no requirement for specifying endogenous and exogenous variables and consequent identifying restrictions like simultaneous equations system models. If there are no contemporaneous terms on the right-hand side of the VAR specification, it is possible to simply use OLS separately on each equation.

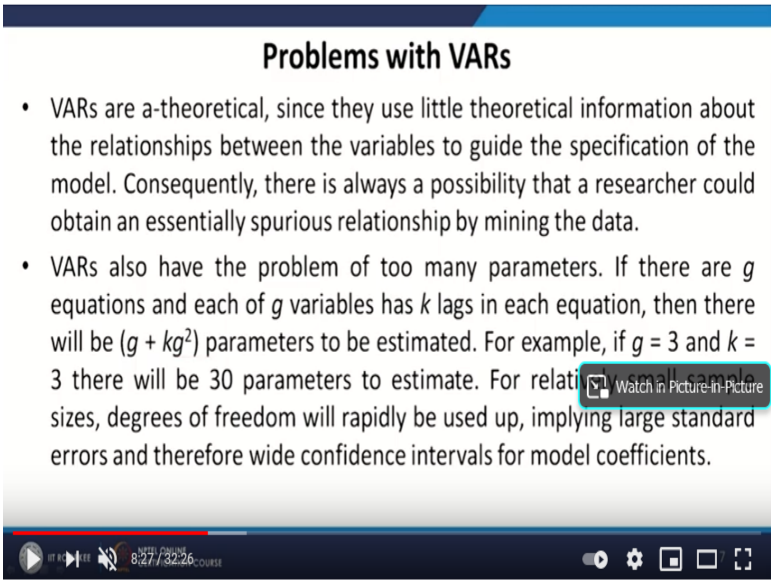
So, the contemporaneous terms here are the terms that refer to the same period, as the left-hand side. So, for example, (refer slide time: 6:05). So, no contemporaneous terms on the right-hand side. And therefore, these equations can be estimated simply using OLS by estimating one after another. So, one equation at a time.

This arises from the fact that all variables on RHS are predetermined, we know the values of them, that is, at time t they are known. This implies that there is no possibility for feedback from any of the LHS variables to any of the RHS variables. So, this implies that we are having all lagged values on the right-hand side, which implies that by the time, time t comes, we have observations on all these variables.

And that is why it is not possible that the left-hand side is explaining the right-hand side. The left-hand side does not explain the right-hand side. So, there is no feedback from the left-hand side to the right-hand side. The right-hand sides are already observed variables, and that is why there is a problem in estimating these equations using OLS.

The forecast generated by VARs is often better than traditional structural models. This could perhaps, arise as a result of the ad hoc nature of the restrictions placed on the structural models to ensure identification. In structural models, that is, simultaneous equation systems we needed some identification restrictions. So, now since we do not need to come up with similar restrictions, this is an added advantage. But then, VARs have certain problems also. So, VARs are a-theoretical, which implies that it is actually not generally driven by any theoretical understanding.

(Refer Slide Time: 8:14)



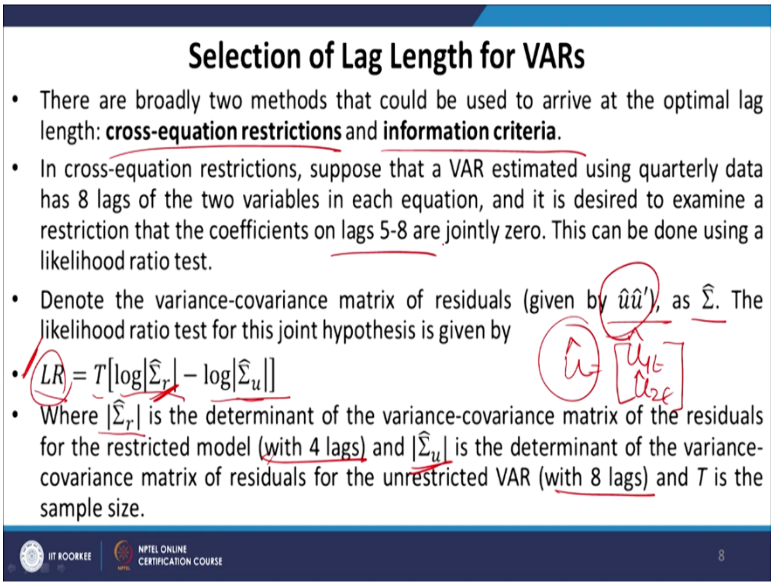
So, since they use little theoretical information about the relationships between the variables to guide the specification of the model, we call them a-theoretical. Consequently, there is always a possibility that a researcher could obtain an essentially spurious relationship by mining the data.

So, since it is not driven by any theoretical understanding, it may happen that I estimate a VAR model involving 2, 3, 4, 5 variables, and I also observe certain relationships between them, but then the relationships are actually spurious in the sense they are not meaningful from economics perspectives. So, because it does not have any underlying theory explaining the relationships between these variables.

VARs also have the problem of too many parameters. If there are g equations and each of g variables has k lags in each equation, then there will be $(g + kg^2)$ parameters to be estimated in the entire system. For example, if $g = 3\, and\, k = 3$, then there will be 30 parameters to estimate. For a relatively small sample size, degrees of freedom will rapidly be used up implying large standard errors and therefore wide confidence intervals for more model coefficients.

So, what is happening here is that even if I work with say, roughly, 50 years of data, but if I have 3 variables in a way or model, then I will be estimating 30 parameters, and left with only 20 degrees of freedom. So, degrees of freedom are being used up very quickly when we have a VAR system with many variables and relatively the sample size is small.

(Refer Slide Time: 10:01)



Now, I come to the point of how to select the lag lengths for VAR models. There are broadly two methods that can be used to arrive at the optimal lag length. The two alternative methods are cross-equation restrictions and information criteria. In cross equation restriction, suppose that a VAR estimated using quarterly data has 8 lags of the two variables in each equation. So, for example, the example that we had taken in the beginning that will have (refer slide time: 10:32).

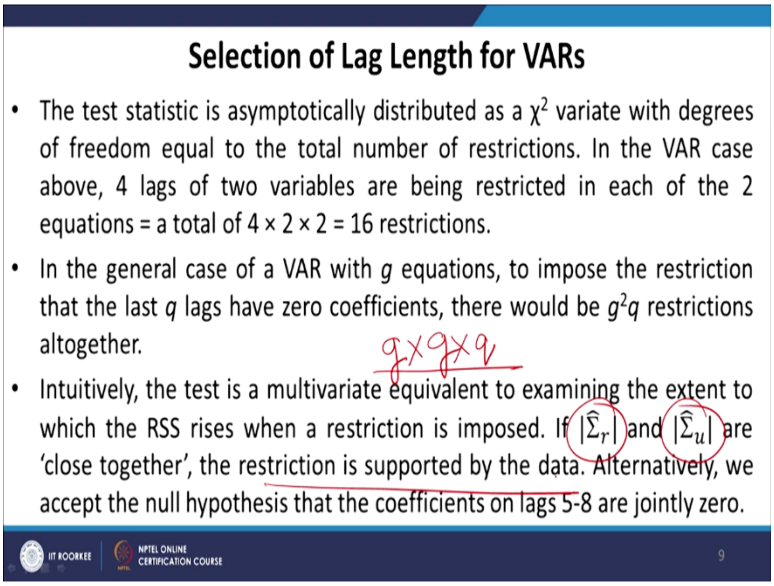
Now, it is desired to examine a restriction that the coefficients on lags 5 to 8 are jointly 0, if we want to find out whether the coefficients of lags 5 to 8 or jointly 0 then what do we need to do this can be done using a likelihood ratio test. So, denote the variance-covariance matrix of the residuals, the residuals are (refer slide time: 11:06- 11:59).

Now, as I have just mentioned that this $\hat{\Sigma}_r$ this sign actually refers to the determinant of the variance-covariance matrix of the residuals for the restricted model, so $\hat{\Sigma}_r$ is actually the covariance variance matrix obtained from the restricted model where $\hat{\Sigma}_u$ is the variance-covariance matrix obtained from the unrestricted model. The unrestricted model has all 8 lags and the restricted model has only 4 lags because we are putting the restrictions in the restricted model that the coefficients of lags 5 to 8 are jointly 0 and T is the sample size.

So, this is very similar to the kind of F-test we have conducted or considered earlier. But the application of the F-test is more problematic here. And the reason is that, if you expand or convert the expression for the VAR model in terms of the error terms, you would find that the error terms are actually serially correlated. So, as a result of which application of F-statistic is not desirable, because the error covariance variance is not the most efficient one. So, we have the OLS estimates, but the

OLS estimates may not be the most efficient ones. So, that is why we prefer to go for a likelihood ratio test.

(Refer Slide Time: 13:30)

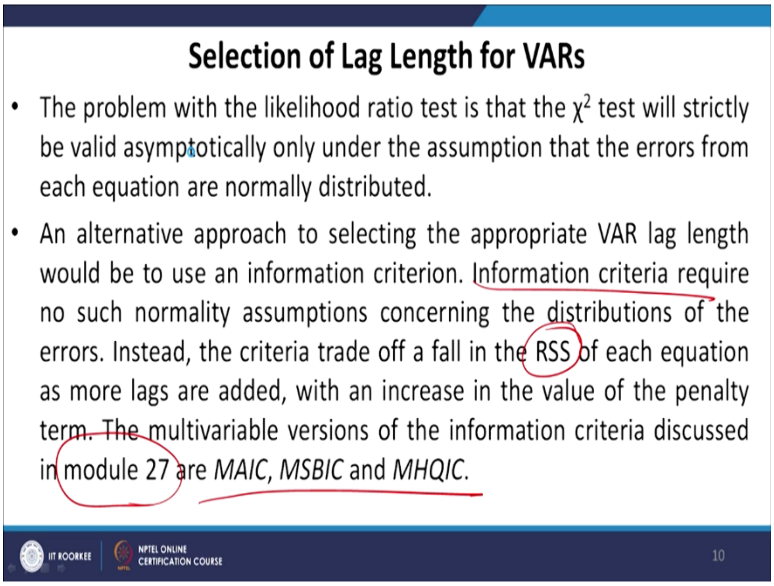


The test statistic is asymptotically distributed as a chi-square variant with degrees of freedom equal to the total number of restrictions. In the VAR case above, 4 lags of two variables are being restricted in each of the 2 equations, which implies that a total of 16 restrictions are there. In the general case of VAR with g equations to impose the restrictions that the last q lags have 0 coefficient, there will be g^2q restrictions altogether. This is because there are g variables.

So, in each equation, I will be having g multiplied by q restrictions and then there are g equations because for each and every variable there is an equation, so that is why I arrive at g^2q. Intuitively, the test is a multivariate equivalent to examine the extent to which the RSS rises when a restriction is imposed. If the determinant of $\hat{\Sigma}_r$ and determinant of $\hat{\Sigma}_u$ are close together, the restriction is supported by the data, which implies that we do not reject the null hypothesis.

And the null hypothesis is that the lags of 5 to 8 lags are jointly 0 or the coefficients associated with the fifth to eighth lags are jointly 0, that null hypothesis is not rejected. And that is why the restriction is supported by the data. Alternatively, we accept the null hypothesis that the coefficients on lags 5 to 8 are jointly 0.

(Refer Slide Time: 15:08)

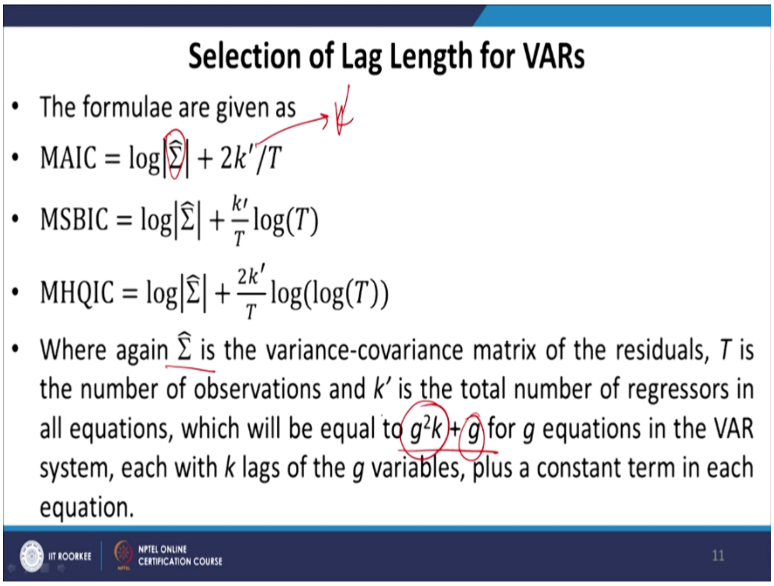


Now talking about the alternative methods of why do you need to consider an alternative method of information criterion. So, the problem with the likelihood ratio test is that the chi-square test will strictly be valid asymptotically only under the assumption that the errors from each equation are normally distributed.

So, an alternative approach to selecting the appropriate VAR lag length would be to use an information criterion. Information criterion was earlier introduced in module 27. Again, in the context of selecting the lag lengths of AR, ARMA models. So, they were in the context of univariate series, and now, we are in the multivariate arena, so we are having multivariate AIC, multivariate SBIC, and multivariate HQIC.

So, information criteria required no such normality assumptions concerning the distributions of the errors. Instead, the criteria, as mentioned earlier explained earlier, trade-off a fall in the RSS of each equation, RSS here stands for a Residual Sum of Square as more lags are added with an increase in the value of the penalty term and that is why they are preferred over the other method of cross equation restrictions at least when we do not expect the data to have or the errors to have a normal distribution.

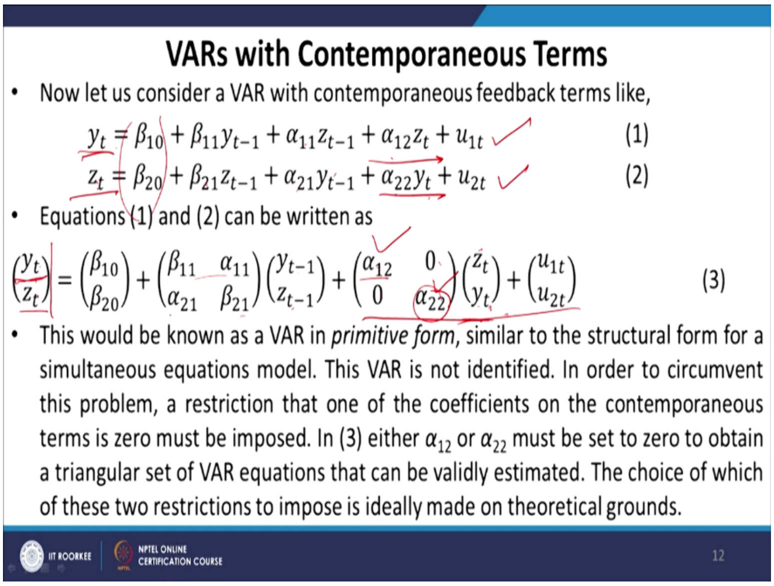
(Refer Slide Time: 16:36)



So, the formulas are very similar. Only in place of the previously estimated error variance, we are having this $\hat{\Sigma}$ and the rest of the expressions are the same. Another difference is that we are having k' here. Now, here, $\hat{\Sigma}$ is, of course, a variance-covariance matrix of the residuals, T is the number of observations and k' is the total number of regressors in all equations, which will be equal to $g^2k + g$ for g equations in the VAR system, each with k lags of the g variables plus a constant term in each equation.

So, because of those constant terms, we are having a plus g here and for k lags in g equations, we are having g^2k. So, that total number of variables or coefficients would be equal to $g^2k + g$ in the information criterion.

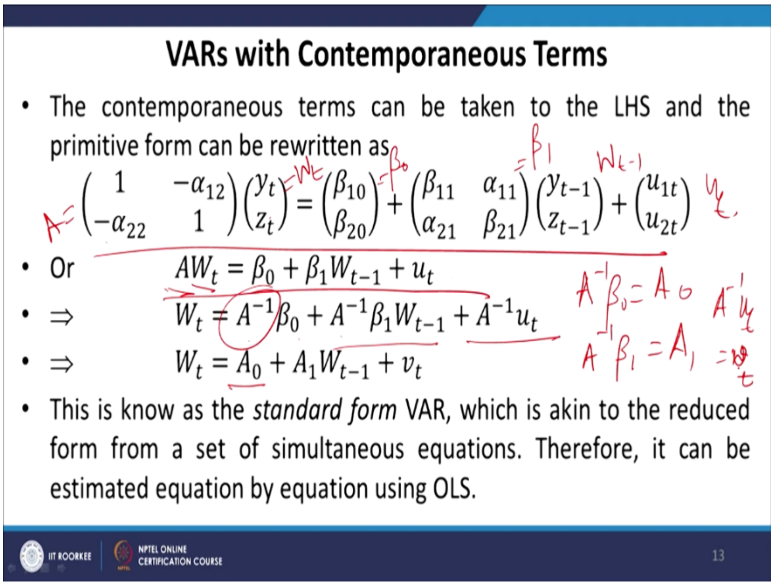
(Refer Slide Time: 17:32)



Now, let us consider VARs with contemporaneous terms. What if we include contemporaneous terms on the right-hand side? So, we have included here (refer slide time: 17:44- 18:28). This would be known as VAR in primitive form, similar to the structural form for a simultaneous equations model. Because you can see that we have endogenous variables both on the right-hand side as well as on the left-hand side. This VAR is not identified.

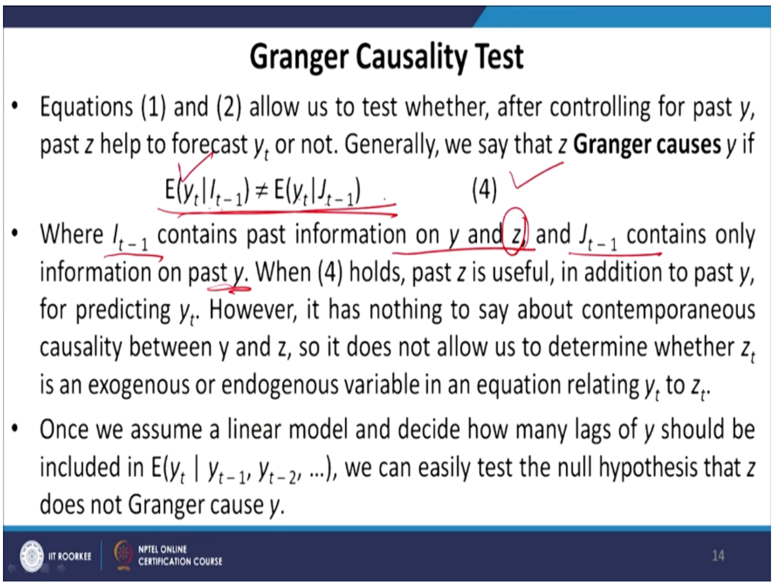
In order to circumvent this problem, a restriction that one of the coefficients on the contemporaneous terms is 0 must be imposed. So, in order to make this system identifiable, we need to have either (refer slide time: 19:00- 19:28). So, we have a recursive system or a triangular system, which was introduced while discussing the simultaneous equation system. So, that is how we can obtain a triangular set of VAR equations that can be validly estimated. The choice of which of these two restrictions to impose is ideally made on theoretical grounds.

(Refer Slide Time: 19:48)



The contemporaneous terms can be taken to the left-hand side and the primitive form can be written as this. So, now we are basically, we have moved (refer slide time: 19:59- 21:05). So, this is known as the standard from VAR, which is akin to the reduced from a set of simultaneous equations. So, we obtain reduced form in a simultaneous equation system in a very similar fashion. Therefore, it can be estimated equation by equation using OLS, the way reduced from can be estimated in a simultaneous equation system using OLS.

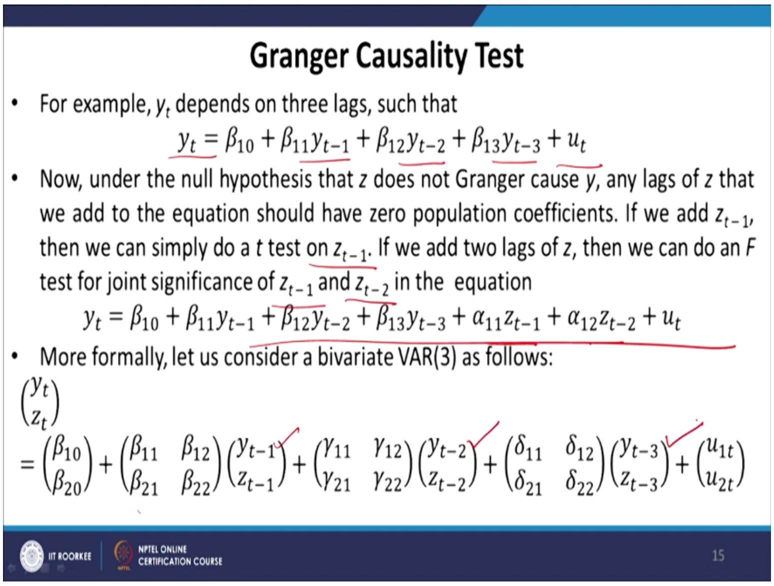
(Refer Slide Time: 21:26)

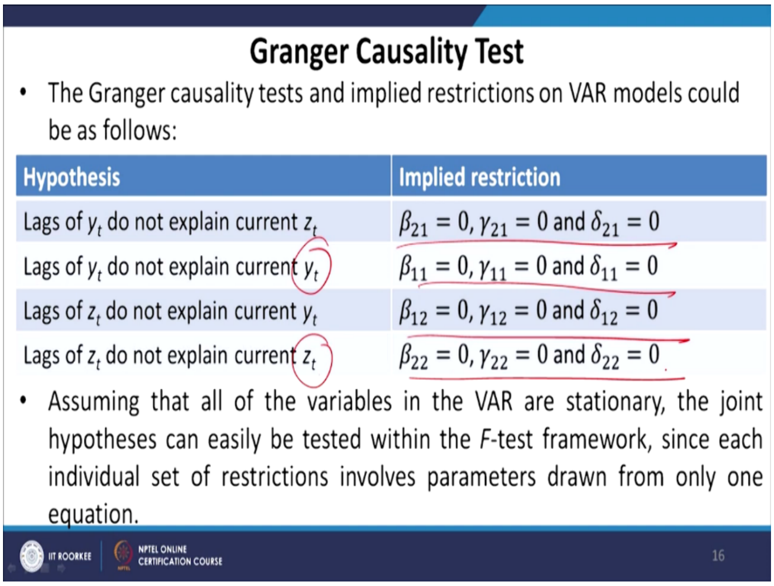


Now, we talk about Granger Causality Test. This is something, which is often associated with VAR systems. So, equations 1 and 2, these two equations, allow us to test whether after controlling for past y past z helped to forecast y_t or not. Generally, we say that, (refer slide time: 21:48- 23:00).

Once we assume a linear model and decide how many lags of y should be included in the expected value of y_t conditional upon its past values, we can easily test the null hypothesis that z does not Granger cause y. Now in this context, let me tell you that the Granger causality test is named after Granger. I think the test was suggested in 1976.

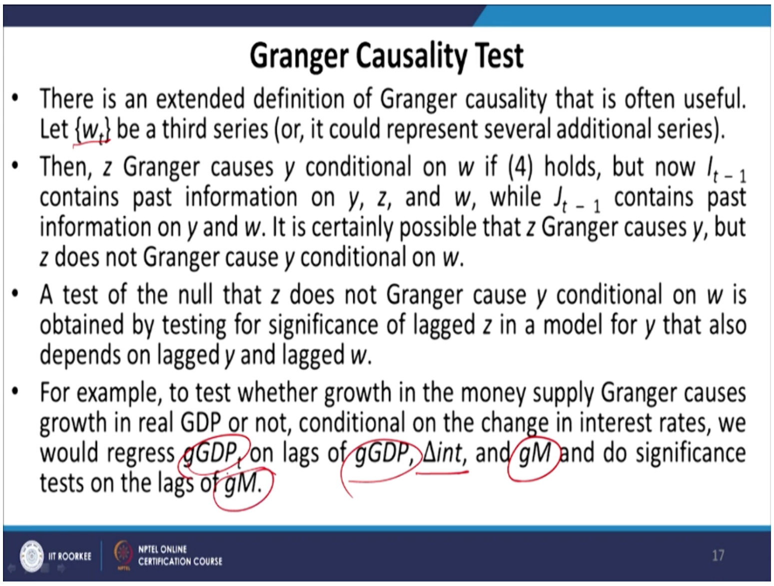
(Refer Slide Time: 23:22)





Now, for example, (refer slide time: 23:24- 25:50). Assuming that all of the variables in the VAR are stationary, the joint hypothesis can easily be tested within the F-test framework. Since each individual set of restrictions involves parameters drawn from only one equation.

(Refer Slide Time: 26:06)

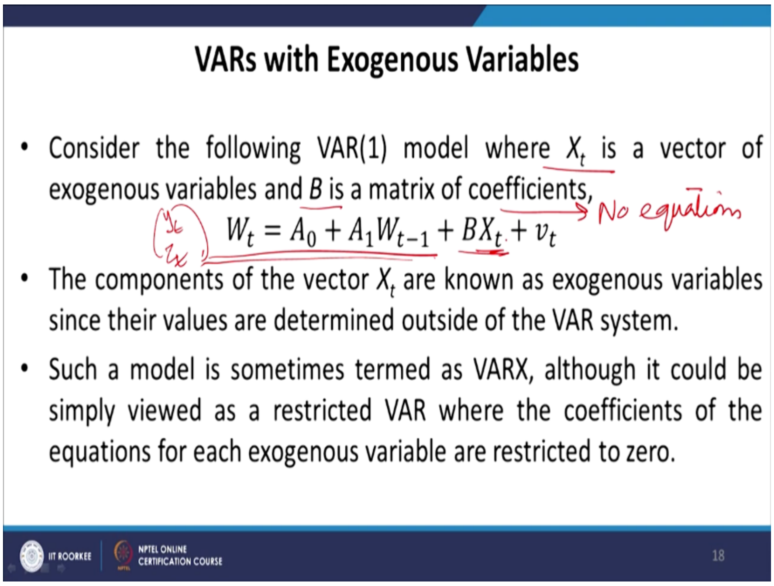


There is an extended definition of the Granger Causality Test that is often useful. So, let us consider w_t be a third series or it could represent several additional series. Then z Granger causes y conditioner on W if 4 holds so, this was my condition 4, and but the thing is that now, I_{t-1} minus one contains past information on not only y and z like previously, but now it contains information on y, z , and w.

While J_{t-1} contains past information on y and w because we are focusing only on z, whether z Granger causes y or not. So, all other variables are there in j. It is certainly possible that z Granger causes y, but that does not Granger causes y conditional on w. Tests of the null hypothesis that z does not Granger cause y conditional w is obtained by testing for significance of lagged z in a model for y that also depends on lagged y and lagged w.

For example, to test whether growth in the money supply Granger causes growth in real GDP or not conditional on the change in interest rates, we would regress $g\ GDP$ that is the growth in GDP on lags of $g\ GDP$, change in the interest rate, and growth in money supply and do a significant test on the lags of only growth in the money supply.

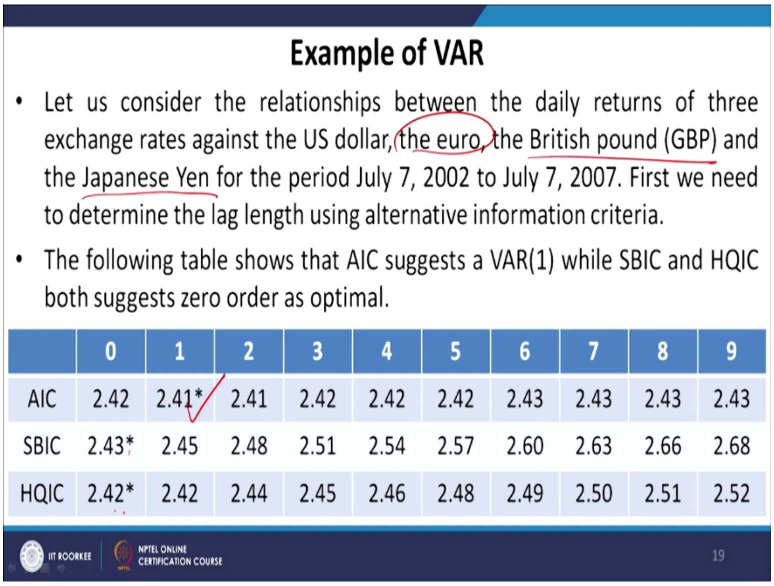
(Refer Slide Time: 27:42)



Now we consider VAR with exogenous variables. This is just a very brief introduction to it. So, we consider the following VAR (1) model while (refer slide time: 27:52- 28:12)

The components of the vector X_t are known as exogenous variables since their values are determined outside the VAR system. And this also implies that we ideally will have no equations for the variables that are included in the vector X_t. Such a model is sometimes termed VARX. Although, it could be simply viewed as a restricted VAR where the coefficients of the equations for each exogenous variable are restricted to 0.

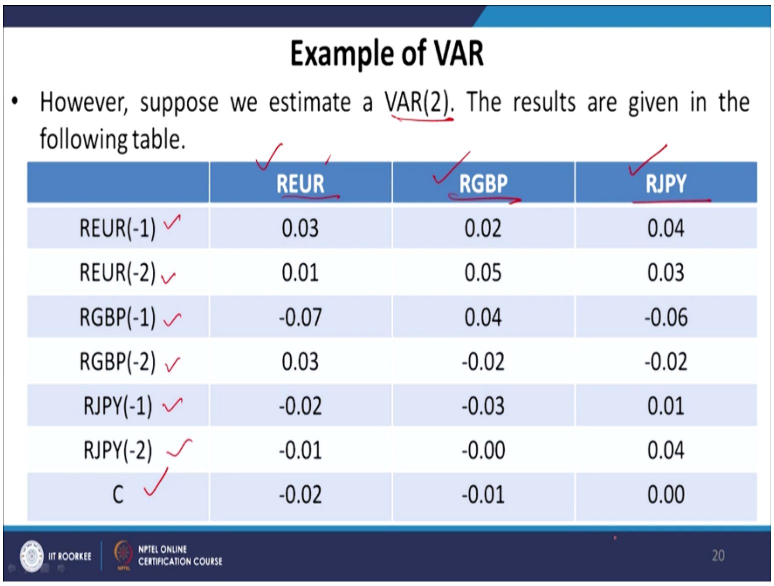
(Refer Slide Time: 28:40)



Now, we take a take an example of a VAR model. How it is estimated or how the estimated results look like? So, let us consider the relationship between the daily returns of three exchange rates against the US dollar that is the Euro the British pound, and the Japanese Yen. And for the period of July 7, 2002, to July 7, 2007, first, we need to determine the lag length using alternative information criteria.

So, the following table shows that AIC suggests a VAR (1) because you can see that all our insignificant a star indicates a significance level adulate rate of 5 percent. So, VAR (1) is suggested by AIC while SBIC and HQIC indicate 0 lag length, that is, they both suggest 0 order as optimal.

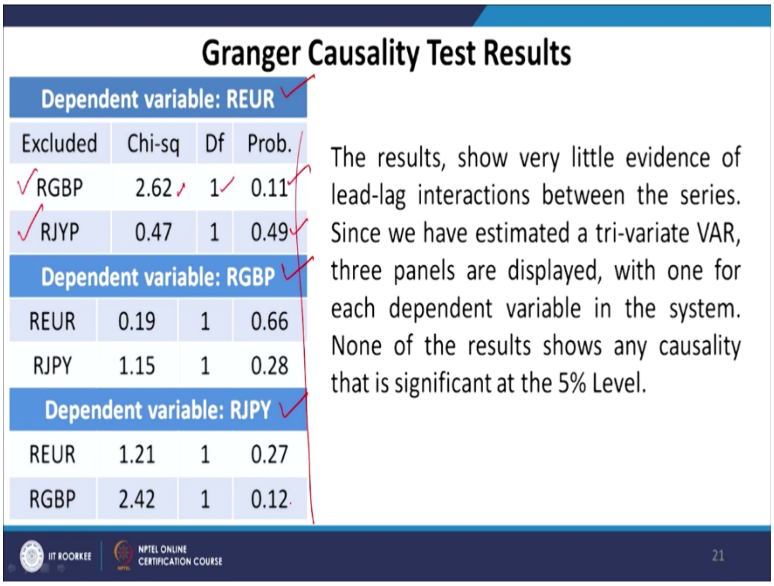
(Refer Slide Time: 29:36)

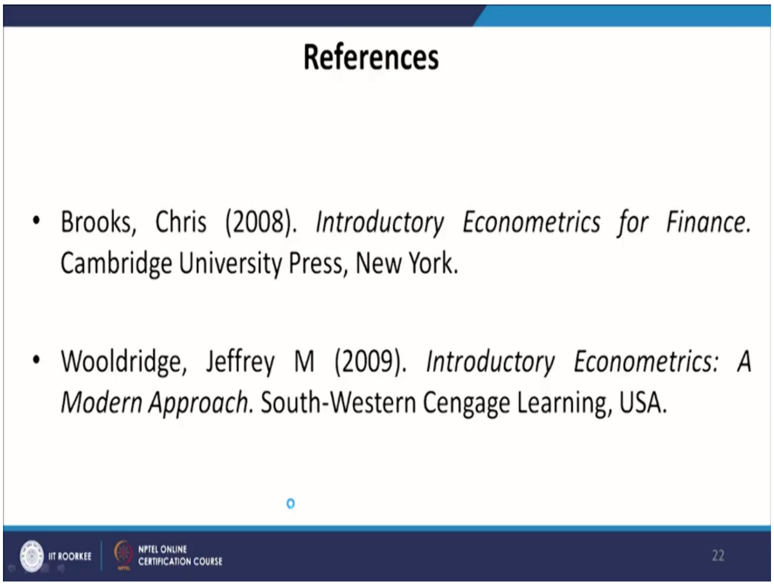


However, suppose we estimate a VAR (2) model, this is just an example. So, the results are given in the following table which shows that these are the variables. So, this is say y_t, this is Z_t and this is N_t. REUR refers to return from euro return from the GBP that is Great Britain pound and this is Japanese Y return from Japanese Yen.

Now, and we have considered VAR(2), so for each and every variable we have two legs you can see and then one constant term. So, this is how VAR results are reported. Now, of course, very, unfortunately, none of these variables appear to be significant, though some of them are probably close to a 10 percent level of significance, but none at 5 percent at least.

(Refer Slide Time: 30:27)





And finally, I also report the Granger Causality Test results. And you can see that, again, the dependent variable REUR here RGBP here and RJPY here. And we are basically excluding RGBP, we are considering the impact of RJPY excluding RJPY, RJYP we are considering the impact of RGBP and things like that. The chi-square values, degrees of freedom, and probabilities are mentioned here.

Again, this shows that none of them none of these are basically significant, which implies that none of the returns actually significantly cause the other returns. The results show very little evidence of lead-lag interactions between the series. Since we have estimated at tri-variate VAR three banners are displayed with one for each dependent variable in the system. None of the results show any causality that is significant at the 5 percent level.

So, that is broadly about the VAR models at all I wanted to talk about. VAR models have some other important applications like calculations of impulse responses, but that would remain outside the scope of this course. Because, again, there is time and space constraint. So that is all about me or model. These are the references I have broadly followed. Thank you.