Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology – Kharagpur

Lecture – 78
Autocorrelation (Contd.)

Hello and welcome back to the lecture on Applied Econometrics. We have been talking about autocorrelation and in the previous lectures we have seen different tests autocorrelation. Now, in this lecture we are actually going to see data, we will be using both stata and R. And we will see how autocorrelation looks like? And how we are going to use different tests to determine the autocorrelation?

And finally we will also talk about how to actually deal with autocorrelation? So, we know that autocorrelation is there and we know it is a problem. But so what how can I actually you know make my model free of auto correlation? So, this would be our topics in this lecture and in next couple of lectures. So, let us start with how can I visualize from data, if autocorrelation is already there or not.

(Video Starts: 01:13) So, let me actually first start with my stata. So, I will be showing you a data set that is about US consumption data set. So, let me import it is an excel format and I need to give it the path, this one table 6 1 excel file. So, let me open it. So now, I can actually visualize this data and I can you know will explain this data set. So, we have year, we have consumption, we have income, we have wealth, we have interest rate and then we have these are the main variables we will be using.

Consumption, income, wealth and interest rate and then we will have this log of consumption, log of disposable personal income, log of wealth and then difference of log of consumption. I will explain this d part later on, d log of disposable personal income, d log of wealth and d of interest rate. So, I will talk about all these different variables essentially the point is we have these 4 different variables which we are going to work with consumption, income, wealth and interest.

Now, for some reason it seems that data for year is not in the right format. So, let me do one thing, let me actually, let us say, let me actually drop this variable. So, now it will be fine. So,

I can see all the variables are at the top of the table. So now, I can write down this command drop here. So, I actually drop this variable because the, for whatever reason the year is not coming in the right format.

So, instead I will create one variable let us say year. So, let us I generate a variable gen year and it is for n and if I do that I will see in my stata. I have actually created this variable, I actually had a variable time I could have used that but no problem. Now, I have created a variable year and it is 1, 2, 3, 4 so, as per the sequence. So, let us say this data is taken from 1951 let us say. So, what I will do? I will simply add year = year + 1950.

So, I will get the values here, assign the values, year = year + 1950. Now, if I go back there so, I will see my year has started from 1951 because I have added 1950 to all of these. Now, what I will do? I will, actually I need to bring the year as the first column because that is the time and I want to see it. So, what I will do is, I will write this command order year. So, this is the basic cosmetics we need to do as a part of the data preprocessing.

So, I got my year here. So, now, I got my you know exactly the kind of table I was looking for. Now, you want to visualize autocorrelation here. So, how do I visualize autocorrelation? So, essentially to visualize autocorrelation I need to get the residual term. So, we know we need to get the u hat term. So, how do I get the u hat term? I have to first define the regression equation. So, I can get u hat term let me actually.

So, let us say you know a very basic regression model I am going to write. I am writing consumption as a function of income and wealth. So, consumption is my dependent variable, income and wealth are my independent variable. So, of course, if my income is increasing it is likely that my consumption will increase and if I have a lot of wealth so, which are already there as stock variable, if I already have a lot of wealth it is likely that my consumption would be higher.

So, in both the cases the variables are you know likely to influence my consumption positively. So, if I just simply write down this equation regress consumption and my independent variable is income and wealth. So, if I do this so, I get a nice regression equation I get R square value it is 0.99. It nothing can be better than this. I just R square value is also pretty good, I have 54 observations.

Now, I have to check, if my model has some problem autocorrelation or not. So, let me actually. So, how we do that? We have to basically look at the u hat. We have to estimate the u hat, the error term and from there we will actually get and if we plot the u hat will actually get you know, if there is any sort of correlation among the error terms. If there is some sort of relationship among the error terms.

So, we will find that you know there is a autocorrelation present in my model. So, how would I do that? So, essentially I would write predict let us say I define a variable u hat and u hat is basically residual. So, stata understand is common resid or residual whatever you write. So, if I write this so, I get the u hat. So, I got my predicted u hat. So, if I plot u hat of course I have to give time.

Its plot u hat against what? So, if I plot u hat with year. So, my time variable here is year. So, then I see that there is a lovely sort of you know graph appearing and I can see that you know it looks like there might be some pattern because it is narrow at the beginning and that is expanding and there could be some sort of you know pattern like. But we have to still see perhaps I should plot a line.

So, how we do it in a stata is 2 way line, I think this is a command line and then we will write u hat and we had year. So, if I write this statistic in this command. So, we will see the line exactly that what we are looking for. We got a beautiful sort of plot where you can see that the error terms of every period are actually sort of there is some sort of relationship. So, every period so, first in 1950, it is gradually declining declining declining till 1970.

And then slightly peak that started picking up then sudden fall and then again some spikes and continuously it is expanding the error time is continuous expanding. So, it is quite likely that there would be some sort of you know correlation autocorrelation structure present in my model. So, there are some correlations in the error terms and there is autocorrelation present in my model.

So that is how we see in stata, if you know this data set how do we visualize you know looking at the residual plot? Now, let us go to R. So, I was actually you know doing this a little bit. So, how we do it in R? So, we will first we need to run the working directory. So, I

got my working directory is already fixed here. So, I do not have to set it then for reading excel file, I need to run this command read excel and I have already loaded this data set in my R.

So, I can just you know give a name U S consumption and that is basically the name of the data set. So, my R program will actually recognize whenever I write US consumption to be the exact data set I have. Because in my R program you can see there are many many files open. So, I need to tell which exact file I am talking about. So, US consumption is the name of the file, I am basically giving this name and if I run it.

So, it has actually recognized the file. Now, I will actually copy some of the codes I was writing. So, let us say I want to again let us say we want to visualize and I think there is some problem with the time variable, year variable. So, again what I am going to do is, I am going to create a variable called time. Just like before and what I will do is, I will let us say or I can actually created this variable. So, let us not just waste time on this part.

So, I actually had this time variable here. So, this time so, I had this 1, 2, 3, 4 up to 54 because I had 54 observation. So, what I am going to do is, I am just simply going to just a second there is too many R files open here. I am simply going to write. So, since it is 1, 2, 3, 4, I want to give it a exact year. So, I can write US consumption time is equal to so, we use an arrow sign in R is equal to US consumption + let us say 1950.

So, if I write that so, essentially it will take the value from 1951. That is what it should do. Now, once I have my year data ready with me. So, what I will do like previously in like we did previously. So, I need to run a regression here also. And if I want to run a regression I need to install a package called L M test. So, I am just copy pasting instead of typing down all the codes. So, if I install this L M test package.

So, this package you need in order to run a regression or doing econometric you know sort of activities. So, once you install it you have to load it and load when you have to load it you need to write library the same L M test. So, R has completed downloading the package. Now, I can run this. So, or else actually you know loaded it. Now, I will run a regression, I will simply run a regression like I did just now using stata.

So, I am just going to write, I am going to name a model let us see it is I have named this as mod, model and in my the model is essentially, I am running a regression equation where my dependent variable is consumption just like before and my 2 independent variable income and wealth. So, if I do that and if I run it, I get my regression equation. I get my regression done. Now, in order to visualize unlike stata you have to actually write a command.

So, if you can write print. So, it will give the summary of the model or you can actually write the summary. It will actually provide you the details of the model. So, let us write down both. So, it has given the coefficients and if you run the summary code. So, then you will get all the coefficients, the p value, standard error, the value of the t statistic, as well as the R square adjust R square and everything that you need in a regression equation.

So, like before you again want to understand how the error terms are related and in R is really simple. So, all you have to do R again like stata, R also understand the residuals. So, plot. So, basically in this case I have to write down the residual of this model. So, essentially I have to write the resid comma resid model command. So, plot resid model and if I plot it I will see something like this, exactly the same data.

So, the of course that this you know the way this points will be scattered will be same and if I want to get a line or something. So, in that case I have to write something like a type. So, plot resid type is equal to o. So, there are different types you just check it in the internet. So, there are you know you can get l, h, s for step function and so forth. So, for different output you can actually sort of give a command to R, what type of you know sort of plot you want to see.

So, here I wanted to see a line plot. So, I have given this command o and I got the how the points are actually connected. So, essentially the same thing we have seen in this stata we can see in R. So, essentially this is how you by plotting the error terms you can actually see, if there is any structure of autocorrelation present in your model and it is advisable before even, if you do the test Durbin Watson Test or Breusch Godfrey Test.

It is advisable that if you suspect that your model has autocorrelation, it is always advisable that you actually plot the residuals to see, if the autocorrelations are present or not. **(Video Ends: 15:00)** So, with this I will end this lecture here and in the next lecture we are actually

going to you know use the same data set to see, if there are Durbin Watson you will basically will run Durbin Watson Test.

We will also run Breusch Godfrey Test. And then we will talk about, how to actually get rid of the autocorrelation problem which we have. Thank you.