Time Series Modelling and Forecasting with Applications in R

Prof. Sudeep Bapat

Shailesh J. Mehta School of Management

Indian Institute of Technology Bombay

Week 08

Lecture 39: Testing for Causality

Hello all, welcome to this course on Time Series Modeling and Forecasting using R ok. Now, again the title that you see in front of you states Causality Tests and Further. So, again just to quickly give you a short overview of where we stand this week. So, the entire focus area in all the previous lectures this week has been co-integration right. So, the main topic which we are discussing so far this week is co-integration.

So, firstly we started this week by explaining what do you mean by co-integration and next when can we say that two series XT and YT are indeed co-integrated or not. And there we argued couple of things that if Xt and Yt are individually integrated with certain order let us say I1 or I2 etc. But if certain linear combination of Xt and Yt exists let us say where beta is the coefficient. Whereas, beta is also called as a cointegration vector, but the other words is that if there exists a certain linear combination of x t and y t such that the linear combination happens to be I 0, then one can actually say that x t and y t are cointegrated. Okay.

And then we talk about a couple of other things such as what do you mean by spurious regression, right? And just in the last lecture, we talked about several tests by using which one can actually find out whether two series of XT and YT are cointegrated or not. Okay. And there, if you remember, we talked about four or five different tests. Let's say Engelgranger or Johansson test, then Phillips-Oliaris, and then ARDL bounce test, etc.,

And all these tests the underlying H naught is that the X t and Y t are not co-integrated while the alternative hypothesis was the H t and Y t are co-integrated basically. And then towards the end of just last lecture we saw a couple of examples where testing of co-integration is sort of important to analyze both the series as further ok. Now, in today's lecture we will try to explore a few ideas about causality tests. So, firstly what do you mean by causality tests and a few other aspects related to causality tests. All right.

So firstly, we have to define what exactly one means by causality test. So the heading is causality tests. So causality tests are statistical methods or tools used to assess whether one time series can predict or cause changes in another. So, this term causality is as the name suggests that one has to find out whether one of the variables let us say xt causes any change in one of the other variables let us say yt or even vice versa. So, if any of the variables xt or yt causes a change in the other variable then we can say that causality between xt and yt exists.

And I am pretty sure that many of you might have heard about this very famous statement in almost all the, that you hear in almost all the statistical courses of regression and so on, that regression or correlation does not imply causation, right. So, even if you have high correlation between Xt and Yt, one cannot say that Xt causes Yt or rather Yt causes Xt, right. So, high correlation does not necessarily imply that there has to be some causation between Xt and Yt. But the focus of the entire today's lecture would be more of causation. So, where one of the variable causes some change in the other or vice versa.

So, just to summarize one more time that causality tests are statistical tools or techniques used to assess whether one time series can predict or cause some changes in the other. So, in time series analysis the most widely used causality test is Granger causality right. So, the most widely used causality test in time series literature is Granger causality which determines if past values of one of the variables x contains information that helps to predict the future values of another variable say y. So, the idea about Granger causality is just to find out if there exists some causality between x t and y t such that the past values of one of the variables let us say x contain some information that helps to predict the future values of another variable let us say y. Okay, so now we will sort of break it down.

So, we will try to understand what exactly does one mean by Granger causality in a bit more detail, okay. So, Granger causality is based on the premise that if variable x Granger causes variable y. So, we have this technical term that if there exists any causality between x and y, we will say that the variable x Granger causes variable y then knowing the past values of x improves the prediction of y beyond any of the information contained in the past values of y alone. So, again let me break this statement down just to make it simplified. So, let us say so you have two variables x and y and assume that x Granger causes variable y. So, let us say each of the series now let me write down x t and y t since we are dealing with time series structure.

So, x t and y t are two different time serieses. And each of the series would have its own past information, right, in form of lag. So, let us say Xt minus 1, Xt minus 2, etc. Similarly, Yt would also have its own past lag. So, Yt minus 1, Yt minus 2, etc.,

So, if it turns out that xt Granger-causes variable yt or there is some causality between xt and yt, then knowing all the past values of x—for example, xt minus 1, xt minus 2—improves the prediction of y. So, remember one thing: the focus here is to predict y, which is again, as usual, part of any regression assumption we make, right? So, xt is the independent variable, and yt is the response, and the goal in regression is to predict y. So, similarly here, assume that the goal is to predict y. So, if there exists some Granger causality or simply causality between xt and yt, then using the past values of xt improves the prediction of y beyond any of the information which might be found in the past values of y itself. So, of course, how do you do forecasting or how do you do prediction?

So, one can actually do forecasting based on its own past values, isn't it? But on the other hand, if you have another variable, let us say xt, which is Granger-causing y or there is some causality between x and y, then one can actually add to the information from the past values of x also. So, importantly, Granger causality is not the same as true causation, by the way. So, Granger causality is exactly not the same as true causation. It only assesses the predictive relationship between the variables, right?

So, how could—or rather, how well—can one predict one of the variables, let us say y, based on its own past values, of course, but also based on the past values of some other variable, let us say x? So, under all these situations, one can say that x Granger-causes y, but again, Granger causality is not the same as true causation between x and y. So, now, slowly, we will turn our attention to the actual mathematical formulation of Granger causality. So, again, to test if x Granger-causes y, we will compare two models. So, the first model is what?

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i Y_{t-i} + \sum_{j=1}^{p} \gamma_j X_{t-j} + \epsilon_t$$

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i Y_{t-i} + \epsilon_t$$

So, the first model is called an unrestricted model, which is nothing but a model of Y that includes both the lag values of Y and X, okay. So, the first one is unrestricted, which is nothing but a model of Y that includes both the lag values of Y and X. And the second one is called the restricted model, which is a model of Y that includes only the lagged values of Y. Now again, just to summarize very quickly, imagine that you have two time series structures, Xt and Yt, and again, as discussed in the last slide, Xt and Yt could have their own past values, which are nothing but lags. So, x t minus 1, x t minus 2, similarly y t minus 1, y t minus 2.

So, in this entire setup, one can actually talk about two particular models. The first one is unrestricted, where a model of y includes both the lagged values of y as well as x, and the second one is restricted, where the model of y includes only the lagged values of y itself. Now, assuming we are testing Granger causality up to a certain lag, let us say P, we will fix some lag up to a certain lag P, and up to that lag P, we will be testing whether Granger causality exists or not. So, again assuming we are testing Granger causality up to a certain lag P, then the unrestricted model happens to be this. So, y t equals some intercept alpha plus the first summation i going from 1 to P beta i into y t minus i plus the second summation j going from 1 to P gamma j into x t minus j plus epsilon t.

Now, again, I am pretty sure that this equation might seem difficult, but understanding each and every part is not that difficult. So firstly, what kind of a model is this? The first model is unrestricted. And again, remember, what do you mean by unrestricted? So, unrestricted means that yt uses its own past values as well as the past values of xt.

And hence, we should have both these values inside the regression, shouldn't we? So one should actually have yt minus i as well as xt minus g since the model is unrestricted. And once you plug both these series or the past values of XT and YT into the regression model, one should actually have those coefficients as well. So, gamma j and beta i. So, rather alpha, beta i's and gamma j's are nothing but the underlying coefficients of this regression structure. And similarly, if the model is restricted, then y t cannot depend on past values of x. So, y t only relies on past values of itself, right.

So, you have a simplified structure of the regression. So, y t equals a single intercept alpha as before, but here we should have a single summation. So, i going from 1 to p, then beta i into y t minus i plus epsilon t, ok. So, again just to summarize, if the model is unrestricted then there should be more than two coefficients. So, alpha beta i's and gamma j's, but if the model is restricted then y t cannot depend on the past values of x t's right.

So, in this case we should have a much more simplified regression structure where alpha and beta i's are the only set of parameters ok. All right. So, how exactly does one carry forward the test, right? So, the test involves the following hypothesis. So, the null hypothesis on H naught is that x does not Granger cause y, that is all the coefficients gamma j's are 0.

Now, again, let me just for a second go back a slide and then elaborate on what this null hypothesis is telling us. So, the null hypothesis or H naught states that x does not Granger-cause y. So, there is no Granger causality between x and y, which means that all the gamma j coefficients should be 0. And why is that? Because again, if you go back a slide, the gamma j coefficients are attached to x t minus j. So, if there is no Granger causality between x and y, then this entire structure should be close to 0, isn't it?

So, the entire structure should be close to 0, which means that all these coefficients gamma j's should be really close to 0. Hopefully, this makes sense because if Granger causality does not exist, there is no question of bringing this x t minus j into the model itself. So, in other words, all these gamma j coefficients should be very, very small or rather close to 0 such that this entire factor sort of disappears. And then, similarly, the alternative hypothesis is that if x Granger-causes y, then there should be at least one gamma j which is not equal to 0 such that the second factor in the previous equation does not disappear. So, in this context, one can actually use an F-test for small samples or a Wald test for larger samples to determine if adding past values of X significantly improves the prediction of Y.

If the F-test yields a statistically significant result, we would reject the null hypothesis, suggesting that X Granger-causes Y. Now, first, a couple of things to remember here as a summary. So, let us say the direction we are kind of picking here is like this. So, we are kind of finding out if X Granger-causes Y and not the other way around. So, I think this first point should be etched in the mind that the direction is from x to y, similar to any other regression equation we have. And in this period, I can actually use two tests.

So, the first one is the F-test, which is a usual test, which is nothing but ANOVA that we used in regression. And then, the F-test could be used for smaller samples, but when the sample size is large, one can actually apply the Wald test. But either way, if one applies either the F-test or the Wald test, how do you conclude that X Granger-causes Y or not?

Just by looking at the value of the test statistic. So, if the value of the test statistic happens to be bigger than the underlying critical value, or if the p-value happens to be less than alpha, then we can say that one can actually reject the null hypothesis and simply state that X Granger-causes Y.

Okay. And whenever the conclusion is that X Granger-causes Y, the interpretation of the statement is that past values of X significantly improve the prediction of Y. Does that make sense so far? All right. So, now we will discuss some assumptions of Granger causality. So, the first assumption is stationarity.

So, both the series XT and YT should be stationary. If not, they must be transformed, often by differencing. So, this is a very important assumption when it comes to testing or checking for Granger causality. So, the stationarity assumption again says that both the series XT and YT should be stationary to start with, and if not, they must be transformed or converted to a stationary series, often by differencing. Then, the second assumption is lag length selection.

So, again, if you remember, once we construct that regression equation, that summation ran from i going from 1 to p and again j going from 1 to p, right? So, we fix the lag length to be p there. So, again, the choice of fixing that lag length, let us say p or some other number, is crucial. And this can be determined by using criteria like AIC and BIC. So, how exactly?

So, probably I will explain in very brief kind of an idea that, let us say, the idea is now to fix a particular lag length. So, here what I can do is, I can fix certain values of P, so let us say 5, 6, 7 up to let us say 10. And then, by fixing all these individual values for the lag length P, I can again run my regressions every time and look at the AIC values or the BIC values. And finally, whichever AIC values are least, I can go with that regression model or that lag length. Make sense so far?

So, this is one step of choosing the lag length because choosing the lag length is also crucial. Now, the third assumption is no cointegration for VAR model. So, if X and Y are cointegrated, then using an ECM. So, again, if you remember what ECM is. So, ECM is error correction model, right?

Then, using an ECM or vector error correction model (VECM) is more appropriate, OK. So, one of the assumptions of Granger causality is that there should not be any co-integration existing between XT and YT. And again, this is more specifically for the

VAR model. So, if X and Y are co-integrated, then instead of using an ECM or a vector ECM (VECM), it is more appropriate rather than checking for Granger causality. So, these are some assumptions under the idea of Granger causality.

So, the first one is stationarity, then appropriate lag length selection, then there should not be any co-integration in the underlying series, and so on. Now, we will talk about a few further extensions of Granger causality, right? So, the first one is VECM causality, or rather, vector error correction model causality, OK? So, what do you mean by that? So, just in the previous slide, we saw that this was kind of a limitation, right?

So, if X and Y are co-integrated, then applying ECM or VECM is slightly more appropriate than checking for Granger causality, OK? So, again, we will try to emphasize a bit more on this idea that if two variables are co-integrated, One can actually use a VECM or a vector ECM. And in this case, the Granger causality test should include both the error correction term, which would give you the long-term relationship, as well as the lag differences of the variables, which would give you the short-term dynamics. So, the moment one finds out that the two series, XT and YT, are indeed co-integrated, then one can actually shift to using a VECM.

And how should you use a VECM? That Granger causality test should include both the error correction term, which would give me the long-term equilibrium relationships, and the lag differences of the variables, which would give me the short-term dynamics. Then, the second extension could be non-linear Granger causality. So, by the way, all these are extensions of a standard Granger causality, right? So, let us say the first one is VECM causality, which assumes that XT and YT are also co-integrated.

So, whenever XT and YT are co-integrated, how do you handle Granger causality, right? Then, the second one could be a slightly different modification to the standard causality, which is non-linear Granger causality. So, again, of course, some relationships may be non-linear, requiring models like neural networks or certain kernel-based methods to assess causality. So, if, let us say, XT and YT are non-linear to start with, then, of course, a standard Granger causality test—or rather, a standard Granger causality between XT and YT—might not be useful. So, in this case, if XT and YT are non-linear, we can actually extend a standard Granger causality to something like a non-linear Granger causality, etcetera.

Then, there could be one more extension to Granger causality, which is called instantaneous causality. Now, what do you mean by instantaneous causality? So, this actually tests if changes in X at time t immediately affect Y at time t, which can be explored within a simultaneous equation model framework. So, by the way, this is a very famous kind of framework called SEM. So, many of you might have heard about the SEM structure.

So, SEM stands for simultaneous equation modeling. But in other words instantaneous causality means that if change in the x immediately affect the other variable which is y at the same time. So, changes in x at time t affect the other variable y at the same time t, right. So, all these changes or all these deviations or all these kind of intricacies should happen instantaneous, right. And such a causality can be more explored within a SEM kind of a framework.

So, again I will say just to summarize that all these are extensions of a standard Granger causality that what or rather how could one handle situations if let us say two series are co-integrated or Xt and Yt are non-linear or there exists some instantaneous kind of changes between Xt and Yt at a particular time t. Okay, so now how do you test for Granger causality? So, now that we have developed a few ideas about what do you mean by Granger causality, right? And a few extension of that. So, the first test is called as the Haugh-Pierce test, right?

Again, it is a very famous test for checking for Granger causality. So, the name is Haugh-Pierce test. So, the Haugh-Pierce test is a statistical test used to assess Granger causality in the context of two stationary time series by examining their cross correlations, right? Now, again some of the important assumptions are underlined here is that the both the series should be stationary to start with. So, one can actually apply this Haugh-Pierce test to test whether X Granger causes Y or not with the assumption that both Xt and Yt are stationary time series by examining their cross correlations.

And such a test, or rather the Haugh-Pierce test, is particularly helpful for identifying if past values of one time series can improve the prediction of the other series. So, the past values of Xt can improve the prediction of the other series, which is Yt, often without the need for estimating a model directly, right? So, this is just a broad summary of what you mean by the Haugh-Pierce test and under what assumptions or under what conditions one can apply the Haugh-Pierce test, right? And so on, okay. All right, but then what exactly is the procedure, right? So, what exactly is the procedure of applying the Haugh-Pierce test?

So, the first step is to pre-whiten each of the series. So, again, we have a technical term here called pre-whiten. So, what do you mean by pre-whiten each series? It means fitting a simple autoregressive integrated moving average, or rather an ARIMA model, to each time series separately. So, let us say if you are working with two time series, Xt and Yt, then what one can do is start by fitting or modeling each of the series individually through an ARIMA kind of model structure, right. So, this step isolates the residuals for each series, effectively removing any autocorrelation in the data.

So, even if there exists some autocorrelation between, let us say, Xt and Yt, which is called cross-correlations by the way. So, once you individually model Xt and Yt using an ARIMA kind of structure, then this step sort of isolates the residuals for each series, effectively removing any autocorrelation in the data. And then the second step is to compute the cross-correlation of the residuals. So, the cross-correlation function, or in short, the CCF function between the residuals of the two series, is then found out or is then calculated. And this CCF represents the relationship between the unexpected or unexplained changes in one of the series and the past values of the other.

So, once you pre-Y-turn or rather once you fit individual ARIMA models in both the series correctly, then the second step is to compute the cross-correlation of the underlying residuals. So, let us say you fit an ARIMA model here on Xt and an ARIMA model on Yt, then you gather both the residuals and compute the cross-correlation of both the residuals. Now, again, just to summarize, the cross-correlation function or the CCF function between the residuals of the fitted ARIMA models of the two series is then calculated. And this CCF represents the relationship between the unexpected or unexplained changes in one of the series and the past values of the other. So, such cross-correlations can actually tell you whether the residuals of one of the ARIMA fits have any sort of relationship with the residuals of the other ARIMA model.

Now, again, in this Haugh-Pierce test, what exactly is H0? So, the null hypothesis is that there is no causality between the two series, meaning the cross-correlations are 0. And the alternative hypothesis or H1 is that there is some causality between the two series, indicating significant cross-correlations. Now, again, as before or rather in all the earlier tests that we have seen in the last lecture, the H0 stays pretty much the same: there is no cointegration or there is no causality between the two series, right. And alternatively, there exists some cointegration or there exists some causality between Xt and Yt, right, and so on, okay.

And now, the actual procedure: once you frame both the hypotheses and so on and so forth, the next thing is to formulate the appropriate test statistic. So, in this case, the test statistic is nothing but a sum of squared cross-correlations up to a specified lag, let us say k, okay. So, let us say you fix some lag k, and the test statistic is nothing but a sum of squared cross-correlations up to that lag k. And how do you choose k? So, k is usually chosen based on the data frequency or practical relevance.

So, through some data frequency approach or checking the practicality of the actual underlying data, one can actually fix the lag K. And how does one find out the test statistic Q? So, the test statistic Q is found as Q is nothing but T summation rho hat K square, where capital T is nothing but the number of observations, and rho hat x y k is the cross-correlation at lag k. So, let us say you fix the lag k, and the rho hat tells you the cross-correlation between x and y, and capital K is the maximum lag of interest. So, once you fix this capital K, small k goes from 1 to capital K, right. So, ideally speaking, the test statistic value Q is not that difficult.

$$Q = T \sum_{k=1}^{K} \hat{\rho}_{XY}(k)^2$$

So, it sort of involves the cross-correlations between X and Y up to a certain lag capital K and is the square of that. And now it so turns out that this test statistic or this statistic follows a chi-square distribution with capital K degrees of freedom. So, this Q happens to have a chi-square distribution with capital K degrees of freedom. Now, what may be the assumptions of this Haugh-Pierce test? So, the first important assumption is stationarity, right.

So, the test assumes that both the time series are stationary, meaning they do not exhibit trends or seasonality, okay. And then the second assumption is model appropriateness, that appropriate ARIMA models are fitted to each series to produce residuals free of autocorrelation, and this step is critical because cross-correlations can only represent Granger causality if the series' own past values have already been accounted for. So, the underlying Haugh-Pierce test also comes with a couple of assumptions which one should kind of check for, right. And now the second procedure is called the Hsiao procedure, right, or so probably the H is silent here, it is called the Hsiao procedure. So, the Hsiao procedure is an extended approach to Granger causality testing designed to overcome some limitations of the traditional Granger causality tests.

And by the way, this was proposed by Cheng Hsiao in 1981, hence the name Hsiao procedure. The procedure is especially useful for lag selection in time series data and offers an adaptable way to determine causal relationships. Now, what exactly is the procedure all about? So, the Hsiao procedure combines Akaike's final prediction error (FPE) criterion with the traditional Granger causality, improving both model selection and testing efficiency. So, the Hsiao procedure is kind of an extension of Granger causality by combining the traditional Granger causality with Akaike's final prediction error in some sense.

Now again, we will not go into detail due to paucity of time, and also it will be much more difficult to handle all these different ideas, but I thought of just presenting this Hsiao procedure to you to show how one can extend standard Granger causality. And now, what exactly is the motivation? So, the Hsiao procedure's advantage is that it systematically determines the best lag structure for both variables, avoiding arbitrary lag length choices. It also uses the FPE criterion, which we saw in the last slide, balancing model fit and complexity to ensure the selected model is both accurate and parsimonious. So, all these different ideas point to the fact that one should handle the entire idea of causality with a bit more detail.

So, whether XT and YT are co-integrated or whether they are nonlinear. So, one cannot use standard Granger causality tests, right? So, one can actually go for nonlinear causality, or how do you handle co-integration, or what do you mean by instantaneous causality, right? And then we talked about the Hsiao procedure, etc., or we talked about the Haugh-Pierce test, right? So, all these ideas point to how you handle causality between any two series, right?

Now, obviously, in the next lecture this week, we will finish this entire idea about co-integration or causality through a practical R session, right?

Thank you.