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 36: Cointegration and Further

Hello all, welcome to this course on time series modeling and forecasting using R. Now, it is the start of a fresh week, and as you can see in front of you, the topic for this week broadly would be co-integration and a few other aspects of co-integration. So, initially, obviously in the first session, we will not go into very much detail, but we will study what exactly you mean by co-integration firstly, and of course, several examples would follow. So, firstly, just a couple of things to mention here is that the idea of cointegration would be sort of continued from last week's idea of multivariate time series processes.

So, again, cointegration means what? So, as the name suggests, cointegration means if you have more than one series and not just a univariate time series. So, let us say if you have bivariate series or trivariate series or multivariate series. Then how exactly can you specify the relationship or the interdependency between all those series? So, can we do better, sort of?

So, rather than simply putting forward a multivariate structure or multivariate model, let us say VAR or VMA or VARMA, can you capture the interdependency using some metric? So, this is exactly what the idea of cointegration means. And this further means that we will try to sort of delve deeper into a few aspects of cointegration, a few examples of cointegration in the subsequent lectures. So, the very first idea is a very famous idea which is slightly interesting also, called spurious regression or a spurious regression. So, what exactly do you mean by spurious regression?

So, I will give you a very simple example to start with. So, I have not specified the example here, but just to tell you a very nice example of spurious regression. So, think of a situation where I can draw a plot. So, let us say x-axis versus y-axis. And the x-axis, the variable on the x-axis, let us say ice cream sales, right?

So, let me write down ice cream sales on the x-axis. And just imagine that on the y-axis, you can actually write down, let us say, the runs scored by Virat Kohli, for example, right? Or any batsman. So, runs scored by Kohli, for example, right? So, again, just imagine this example.

So, on the x-axis, what do we have? So, we have ice cream sales per month or per year, whatever, and similarly on the y-axis, we have runs scored by Kohli. So, as they stand, both variables have nothing to do with each other, right? Because ice cream sales have nothing to do with how many runs Virat Kohli will score in that month or that year, etc. But even if you try to plot using a scatter plot kind of thing, you might see that there would be a positive correlation.

So, it might turn out that if you try to obtain a scatter plot between X and Y, it might turn out to be quite linear with a high correlation value. So, such correlations or such dependencies between two unrelated variables are called spurious relationships or spurious regressions, right? Even though X and Y are not related at all, you can still see that the correlation would be, let us say, 0.9, or one can actually have a very nice regression line going through all the points, etcetera. But such regressions are kind of unnecessary or they do not make sense in a way, right, and hence the name is spurious regression. Does that make sense, okay? So, before we delve deeper into what exactly you mean by cointegration, a couple of simpler notations, right.

So, let us say I 0. So, the first notation is I 0. So, I within bracket 0 process. What do you mean by I 0? So, I 0 simply stands for a stationary series without a trend.

So, I 0 process means a simple stationary series without a trend. Similarly, how about I 1? So, I 1 process means an integrated series of order 1. So, there would be some trend, of course, because the underlying process is integrated. But one can actually difference the series once to obtain an I 0 process, right.

So, the essential part here is that even though you have some trend and the series is integrated, but the moment you difference the series once, the one can actually obtain a stationary series or rather I0 process. Make sense? So, slightly different notation from all the previous notations we have seen so far, right. So, I0 and I1. So, now the thing is that modeling two or more time series requires all variables to be I0.

So, again this is one criteria before you enter into forecasting, right. So, before you enter into forecasting, you actually ensure that all the individual series are stationary using some transformation, differencing, whatever. But then do the regression results still hold if some or all the series are I1 instead. So, let us say if you take an example of two different series which are I1. So, do all the regression results still hold or are the underlying regression spurious? this is the question.

So, in this case you can very quickly see that one may face some problems or rather one can actually end up getting some spurious integrations there ok. So, now, the very first example is let us say consider two independent I 1 processes right. So, consider two independent I 1 processes now again very quick refresher is that I 1 means what? So, I 1 means there is some trend And the series is integrated.

But the moment you difference the series once, you obtain a stationary series or I0 process. So, consider two independent I1 processes generated as follows. So, the first series is Xt. And again here, if you notice, how are we generating Xt? So, Xt is nothing but Xt minus 1 plus Vt, where Vt is a pure white noise with mean 0 and variance 1.

$$x_t = x_{t-1} + v_t, \;\; v_t \sim WN(0,1)$$

So, here immediately can you see that x t is integrated or x t is I 1 and how exactly. So, let us say if I bring this x t minus 1 on the left side. Just for a second, if you bring this xt minus 1 on the left side, what will you get? So, if I bring xt minus 1 on the left side, I will have xt minus xt minus 1 on the left side, which is nothing but nabla xt, is it? So, if you bring in xt minus 1 on the left side, I will simply get nabla of xt on the left side and on the right side, I have a pure white noise process.

So, nabla of xt is a white noise or in other words, nabla of xt is stationary, which means that xt is I1. Make sense so far? Similarly, y t. So, again if you bring this y t minus 1 on the left side, again the same thing. So, I will obtain nabla y t on the left side, right. And nabla y t would be exactly equal to some white noise term which is u t, which means that y t is integrated or y t is I 1.

$$y_t = y_{t-1} + u_t, \;\; u_t \sim WN(0,1)$$

Make sense? So, now that we have generated the actual models, the next thing is to simulate some data. So, 10,000 observations are simulated from both the series x t and y t, and these are the plots. So, again, the important thing here is that if you look closely at the plots, what can you see? Or can you tell something about the correlation? Pretty much.

So, I can see that Xt and Yt are sort of negatively correlated, isn't it? So, as Xt drops, at the same time Yt sort of rises, right? So, even though they are completely independent I1 processes, since they are generated using an I1 technique, you can still see that there is some high negative correlation between Xt and Yt. So, such regressions are called spurious regressions. So, even though they are generated from independent families, the ultimate correlation structure or the ultimate relationship between X t and Y t is highly negative.

And here, one can take a step further and actually fit a regression where y is the dependent variable and x is the independent variable. So, a regression of y on x gives the below output, right? So, again, this is a very typical regression output that you see coming from any software. So, be it R or Python, whatever, right? So, the intercept x, the first column, tells you the estimate.

So, the intercept value is 27.91, and then the slope coefficient is minus 1.21, but A couple of things to note here is that the p-value. So, the p-value is highly negative or rather highly small, right? So, less than 2. So, by the way, this notation means 2 into e to the power minus 16 or rather 2 into 10 to the power minus 16.

So, this is a scientific notation that the software outputs, and this simply means 2 into 10 to the power minus 16, right? 2 into 10 to the power minus 16. This number tells you that we have a very, very, very minute p-value or a very small p-value, which sort of tells you, or the conclusion would be, that one will reject the null hypothesis, and since you are going with the alternative, it means that there is some significant relationship between x and y. Make sense? So, this is the first indication, and similarly, if you look at either the R-squared value or the adjusted R-squared value, they are pretty decent. So, roughly around 56 percent each.

So, again, just to summarize. So, even though Xt and Yt were generated using independent families, but still, you can see some essence of regression building between X and Y. So, such regressions are called as spurious regressions. Make sense? So, another example, so let us say again we are generating xt and yt, but slightly differently. So, what exactly is xt?

So, xt is nothing but 0.6 it plus vt. So, vt is a standard white noise, but what is it? So, it is nothing but this guy. So, it is it minus 1 plus wt. So, in short, my It process is nothing but I1, right?

So, can you see a similarity between this It and the earlier Xt in the previous slide, right? So, earlier we generated Xt as an I1 process, but now this It is, in fact, an I1 process. And there is some relationship between this Xt and the I1 process using this constant, which is 0.6. So, Xt is 0.6 into a particular I1 process plus a white noise. Similarly, Yt is nothing but 0.6 into the same It plus another white noise.

So, here we have a small note also. So, note that It is an I1 process, clearly. So, hence the processes Xt and Yt thus involve a common stochastic trend It. So, what exactly is the common part between Xt and Yt, which is nothing but It here. So, It is coming into the picture in both Xt and Yt.

So, in short I1 process which is nothing but IT is the common stochastic trend between XT and YT. So, we see that what happens if you generate X t and Y t using this process now right. So, again as before 10000 observations were simulated from each series X t and Y t and below are the visual plots which again indicate a high positive correlation among the two series. So, can you see that both the series kind of repeat themselves or they have the exact same pattern sort of right. So, on the left hand side you have x t on the right hand side you have y t and as in when x t rises or drops similarly exactly y t either rises or drops.

So, the below plots kind of indicate a very high positive correlation among the two series. Now, let us see that what happens if you run a regression. So, again clearly the regression of y on x gives the below output. The significant relationship clearly points to a question that is it because of the common stochastic trend between the two. So, now the whole question is this is not an example of a spurious regression by the way.

So, now the question is slightly different. So, now the question is that Does this significant relationship? So, again clearly the relationship is significant both visually and through these both tables and why exactly? Again p value is very small.

So, less than 2 into 10 to the power minus 16 right and both the r square values and the adjusted r square values are pretty high. So, almost 1 right. So, 99 percent which means that the underlying regression between y x is pretty strong. Now, the whole question is that since the regression is strong, so is that because of the common stochastic trend which is IT between the two, right? So, this is a very evident question that anybody can ask, right?

So, since the regression between Y and X is throwing some really positive results that there is a very highly significant relationship between Y and X, but then does this relationship sort of come from that common stochastic trend which is IT between the two, ok? Now, further so since XT and YT are I1 processes right. So, clearly since XT and YT contain that common IT process XT and YT are also I1 right. So, if a time series process contains an I1 process as a part of it then the underlying time series process also becomes I1 clearly. So, since X t and Y t themselves are I 1 processes to remove the stochastic trend we require some differencing is not it.

So, if a process is I 1. So, how do you convert I 1 to I 0 by suitable differencing. So, what we do here is we take the first difference of both the series and again run a regression and below is the output. Now, what do you see? So, this is the intercept, this is nabla of x by the way, this is not x or y, this is nabla of x which is the first difference.

And here again, you see that nabla of x is highly significant, again due to the very small p-value. But here, you see something strange or rather interesting: both the R-squared values and the adjusted R-squared values are small, so close to 0, which sort of tells you that the relationship might not be that strong, basically, right? So, one can actually know that the relationship is not that significant here, all right? Okay? But what happens is, however, if you sort of transform the variable slightly. So, now what I will do is I will take this transformation.

So, y_t minus x_t divided by 0.6. Now, here again, why 0.6? Because again, if you remember a couple of slides back, the constant attached to both x_t and y_t was 0.6. Do you remember that slide? Just a couple of slides back, right?

And hence, what I can do is I can inverse transform that slightly. So, what I will do now is that I will take this transformation or rather this model. So, y_t minus x_t divided by 0.6 equals u_t star, and u_t star is another white noise where the mean is 0 but the variance is not 1. So, the variance becomes 1 plus 0.6 raised to the power minus 2. Thus, the whole idea of doing this process is, although x_t and y_t are I(1), right?

So, again remember x t and y t are I 1 right to start with. So, although x t and y t are I 1, there exists a vector beta which is 1 comma minus of 0.6 inverse So, there exists a vector beta. So, try to understand this carefully. So, if you are if you are not very convinced then you should kind of redo this entire lecture one more time or maybe pause the video or something like that right.

So, the important idea here is that although x t and y t are I 1 there exists a vector beta which is 1 comma minus 0.6 to the power minus 1 such that if you pre multiply by beta transpose and then y t x t This transformed variable or this collection happens to be I0 basically. And such a process or such processes are called as co-integrated right. So, the idea of co-integration is what? So, we will try to again explain in subsequent slides, but the idea is that even though individual series is.

$$y_t - \frac{x_t}{0.6} = u_t^*, \;\; u_t^* \sim WN(0,\ 1 + 0.6^{-2}).$$

Thus, although x_t, y_t are $\mathrm{I}(1)$, there exists a $\beta = (1,\ -(0.6)^{-1})$ such that,

$$\beta' \begin{pmatrix} y_t \\ x_t \end{pmatrix} \sim I(0)$$

So, let us say X t and Y t in this case are I 1 or they are integrated, but some transformation of X t and Y t using some vector becomes I 0. So, such processes are called as co-integrated and specifically the vector beta is called as a co-integration vector here. So, hopefully the idea of cointegration is clear now that even though if you start with couple of I 1 processor or couple of I 1 time series. So, let us say x t and y t right, but if you are able to find out a certain vector let us say beta in this case such that the transformed vector or the transformed collection. So, beta transpose into y t x t becomes I 0 then the process X t and Y t are called as co-integrated right and moreover this collection of constants.

So, 1 comma minus 0.6 inverse which is named as beta is called as the co-integration vector right. So, thus it is very important to develop statistical tools suited for capturing the relations between non-stationary time series properly. So, here couple of things are important. So, firstly why non-stationary? So, non-stationary because Xt and Yt were I1 to start with in the previous illustration right.

So, both so, if you look at both X t and Y t they are I 1 and then I 1 is clearly non-stationary right because you have some trend component and then you would want to difference them once to make them stationary right. So, initially X t and Y t are clearly non-stationary, but it is all the more important to develop some statistical tools suited for capturing the interdependencies between such non-stationary time series properly. And the other idea is to overcome spurious relations problem. So, to overcome spurious regressions or spurious relations, instead of using the original series, let us say either Xt or Yt, the series should be suitably transformed so that they can be considered as realizations of weakly stationary processes. Now, again let me let me repeat this statement one more time.

So, even though the statement seems easy to understand, but then what exactly is going on here that to overcome this spurious relations problem instead of using the original series. So, let us say either X t or Y t as they are the series should be suitably transformed and then how exactly. So, in the in the last slide we saw we saw that how did we transform the series by attaching that beta vector to y t and x t right. So, beta transpose multiplied by y t and x t right. So, the collection or the series should be suitably transformed, so that they can be considered as realizations of weakly stationary process.

Because what happens is, once you transform using this beta transpose multiplied by y t x t, the resultant series happens to be I 0, and I 0 is not non-stationary, right? So, I 0 is a stationary process. Make sense? So, just to summarize, in the previous example, the transformation beta transpose multiplied by y t x t leads to a stationary process, which could be analyzed subsequently or as usual. Make sense so far?

So, the idea of cointegration is kind of powerful in that sense—once you identify that a couple of vectors are cointegrated, right? So, this itself tells you how to deal with both of them separately or jointly so as to make them I 0 and analyze them further for, let us say, modeling or forecasting, etc. So, just a summary kind of thing: when the linear combination of two I 1 processes becomes an I 0 process, right? Again, when the linear combination of two I 1 processes— So, X t and Y t in our case, right? And the linear combination is what? So, the linear combination is beta prime multiplied by Y t X t, right.

So, this structure is nothing but a linear combination. So, when the linear combination of two I 1 processes becomes an I 0 process, then we say that the two series are co-integrated. So, X t and Y t will be co-integrated if, even though they are integrated with an order of 1 (I 1) individually, a linear combination becomes I 0. And again, we will quickly kind of go over why we are basically caring about this co-integration at all, right? I mean, why do we care about co-integration?

So, all these are the subsequent points. So, the first point is co-integration implies the existence of a long-run equilibrium. And then, can you sense that? So, if X_t and Y_t are co-integrated to a certain extent, then some linear combination of X_t and Y_t would be I(0). So, in the long run, there will be some equilibrium between X_t and Y_t, is it not?

So, in the long run, one can actually hope that X_t and Y_t would sort of converge or there will be some long-run equilibrium between the two series. Secondly, as discussed earlier, cointegration implies some common stochastic trend between X_T and Y_T, right? And that common stochastic trend was nothing but that I(1) process. So, again, if you remember how we generated both X_T and Y_T a couple of slides back, there was this I(1) process, which was, of course, the common stochastic trend in both X_T and Y_T. So, the second point is cointegration implies a common stochastic trend, which was I_T in our case. Thirdly, with cointegration, we can actually separate short-run and long-run relationships among the variables.

So, if a couple of variables are cointegrated, we can actually separate the short-term relationships between X_T and Y_T and the long-term relationships between X_T and Y_T, okay? Next, cointegration can be used to improve long-run forecast accuracy as well. So, initially, if you know that a couple of series are integrated or there is some cointegration between X_T and Y_T, then this idea can be used to improve long-run forecast accuracy. And lastly, cointegration implies restrictions on the parameters, and proper accounting of these restrictions could improve estimation efficiency, all right. So, in general, if we try to find out whether X_T and Y_T are cointegrated and if they are, then one can actually take some advantage of that in, let us say, modeling, forecasting, or estimation, etc., all right.

So, again, I think this slide kind of focuses on a multivariate angle of the same thing. So, summary for a multivariate situation. So, now that we have seen that, what do you mean by two series which are co-integrated? So, how do you sort of make that statement more general, or how do you extend from two series being co-integrated to more than two? So, summary for a multivariate situation, and of course, there will be some extra notations here, since we are dealing with more than 2 vectors now or more than 2 series, we require some additional or extra vectors or notations here.

So, the first thing is the elements of a k-dimensional vector y are co-integrated. Now, again, how many vectors are we considering here? We are considering k vectors. So, y is nothing but a k-dimensional vector, right? And then again, the idea is, when do we say that they are co-integrated? So, the elements of a k-dimensional vector y are co-integrated of order d, c, right?

So, here we have a pair of orders. So, order D, C, if all the elements of Y are ID. Now, again, just remember, what do you mean by ID? So, ID means that both the series, or rather all the series, are integrated with an order D. So, you basically require differencing d times to make them I(0). So, this k-dimensional vector y are co-integrated of order d, c, d, c, if all the elements of y are I(d), and there exists at least one non-trivial linear combination z consisting of all these variables, right?

$$\beta' y_t = z_t \sim I(d-c)$$

There exists one non-trivial linear combination z of all these variables, right. which is integrated with an order of d minus c. And, of course, here we require that d should be obviously bigger than or equal to c, which should be positive. So, again, just to summarize what is going on here is that we require that there should be some linear combination upon all the individual k vectors which are there in y. So, something like this. So, beta prime multiplied by y t is a linear combination which is termed as z t, and what should the integration of the z t be? So, z t would be integrated with an order of d minus c, basically.

And again, similar to before, beta is called the cointegration vector. And again, the number of linearly independent cointegration vectors is called the cointegration rank. So, I think the last comment is also important that now, obviously, one can actually get multiple such linear combinations also, right? I mean, beta need not be unique. Can you sort of sense that if you are putting forward one linear combination, of course, it has to be non-trivial. So, again, non-trivial means that all the values are not 0s. This is meant by non-trivial, right. So, there should be some non-zero values in that vector also, right.

But again, coming back to my point is that one can actually find out more than one such linear combinations, right? Is it not? Or, in other words, this beta prime need not be unique to a particular construction. So, I can find out beta prime 1, beta prime 2, right. So, let us say you find out beta 1 prime, beta 2 prime, beta 3 prime, etcetera, right. So, what are all these? So, all these are some linear combinations which are termed as corresponding z 1, z 2, z 3, such that this condition still holds true, ok.

So, all these betas are individually called as cointegration vectors and then the number of such vectors the number of such linearly independent cointegration vectors is called as a cointegration rank of that particular situation. Make sense? So, I will say that this is a one slider thing that tells you the transition between two cointegrated vectors to more than two. And of course, since you are dealing with more than two, then of course, we require some extra notations here and there. So, k-dimensional vector, then cointegrated of order d comma c. So, here what do you mean by d comma c?

So, if all the elements of y are id, but there exists at least one linear combination z such that z happens to be integrated with an order of d minus c. So, now we will look at couple of applications. So, where exactly can you see some applications of co-integration and how do you apply that? So, again as seen before the first example is pairs trading in finance. So, co-integration is used in pairs trading strategies where two or more assets are expected to move together over time.

I mean since two assets would be co-integrated one can actually hope that they would be moving together over time. So, even though the individual prices may be non-stationary their difference or for that matter some linear combination would be stationary right. So, this is exactly what we discussed a short while back. So, even though individually they are non-stationary, but some linear combination applied on both the assets would be stationary or I 0. So, how do traders kind of utilize this idea to trade?

So, traders look for pairs of stocks or other financial instruments that are co-integrated and exploit this short term deviations from this long term equilibrium. So, again just a couple of slides back we have seen that if there exists some cointegration between Xt and Yt, there should also exist some long term equilibrium. So, in the longer run Xt and Yt should converge right. And then what happens in shorter term? So, in shorter term they could deviate.

So, the first asset could deviate from its mean or the corresponding mean than the second asset. So, the first asset could go above the mean let us say and then the second asset could drop below the mean right. So, in shorter terms one can actually capture the deviations because one kind of ensures or one can actually hope that in the longer run they would come together again right. So, traders actually look for these short term deviations from this long term equilibrium. So, again when one stock outperforms the other temporarily they may take a long position in the underperformer and short the outperformer, expecting that the prices would converge again.

So, as discussed, if you have one underperformer and the other is outer performer, so one can actually go with a long short strategy in either or, expecting that the prices would again converge in the longer run. So, small example could be let us say stocks of two companies in the same industry. So, let us say HDFC Bank and ICICI often exhibit co-integration due to the similar economic influences right. And then the second example could be something like in energy market. So, prices of energy commodities such as oil, gas, electricity are often non-stationary, but can be co-integrated due to market fundamentals like supply chains, substitution effects or shared inputs right.

For example, oil prices influence gasoline and jet fuel. So, one can actually expect that there should be some interrelationship between all these energy commodities. So, cointegration analysis is used in energy markets to understand the relationships between different commodities or between the spot and futures prices, helping in forecasting, hedging, and creating some arbitrage strategies. And eventually, the relationship between crude oil and natural gas prices might show cointegration, especially in regions where gas is a byproduct of oil extraction. So, broadly speaking, cointegration means that two different series—or, of course, more than two—have some connections between them, and given those connections, in the longer run, they kind of have that long-term equilibrium.

So, in subsequent lectures, we will delve deeper into the idea of cointegration. We will bring in a few more ideas, a few more examples, and, of course, towards the end, we will have a practical session combining all the theory that we have studied this week.

Thank you.