Time Series Modelling and Forecasting with Applications in R

Prof. Sudeep Bapat

Shailesh J. Mehta School of Management

Indian Institute of Technology Bombay

Week 07

Lecture 32: Cross-covariance and Cross-correlation

Hello all, welcome to this course on time series modeling and forecasting using R. Now, again, just to give you a very quick overview of where we are standing this week. So, this would be the second session this week. And the idea we are focusing on this week is multivariate time series analysis or multivariate time series processes. Last week, we talked about some elementary ideas from, let us say, matrix algebra.

And then we tried to put forward some motivation as to why you would want to transition from univariate to multivariate. And hopefully, all the things should be clear to you before you proceed. So before you watch this video, I strongly suggest that if you are not very comfortable with, let us say, matrix algebra results that we talked about towards the end of the last lecture, then you should probably go back and try refreshing some of the very basic linear algebra or matrix algebra ideas. Let us say, how do you define a matrix, then some very basic matrices, symmetric matrix, skew-symmetric matrix, determinants, invariants. So all these basic ideas one should actually refresh.

Now, towards the end of last week, if you remember, we talked about the idea of a random vector towards the end of the last session. We tried to put forward the idea of a random vector. Now again, just to sort of explain what you mean by a random vector one more time, and then we will go ahead with some other notations and some other properties of a random vector, etc. So, a random vector typically is nothing but, let us say, a collection of several random variables. So, let us say x1, x2 up to xp.

So, each one of these is a random variable, by the way. So, x1 is a random variable, x2 is a random variable, xp is a random variable, and hence, a collection of such p random variables would be called a random vector. So, how do you visualize a random vector in a time series framework? So, when you try to analyze a time series, let us say yt. So, yt itself is a random vector.

And why? Because yt, yt plus 1, yt plus 2, etc. Let us say up to yt plus n, and then you have a collection like that. So, each one of the elements in this entire vector is a random variable. I will give you a simple example.

Let us say you want to analyze the stock price of Reliance. So, Yt is the current stock price of Reliance. Yt plus 1 would be one time step ahead. Yt plus 2 would be two time steps ahead. Yt plus n would be n time steps ahead.

But again, of course, each one of those individual prices—so Yt plus 1, Yt plus 2, Yt plus n—what do they denote? They denote the realized stock price on those time steps. Right. And then, can you imagine or can you assume that these are also random variables? The answer is yes.

Right. Because Yt plus 1, Yt plus 2 are random prices. I mean, the price can either go up or go down. Right. Nobody—nobody can tell that for sure.

Right. So, each element—each individual element in this entire collection—is, in fact, a random variable. And hence, this entire thing becomes a random vector. So, one can actually denote this random vector by writing down such a notation. So, Yt inside curly brackets denotes a random vector.

So, anyway, now coming back to the end of the last lecture, if you remember, we talked about the idea of a random vector, and then we sort of discussed some basic properties of a random vector. So, let us see how you define the mean of a random vector, the variance of a random vector, and the covariance between elements of a random vector. So, now we will try to extend that idea. So, the mean vector and the variance-covariance matrix. Now, of course, we discussed this last time, I guess.

So, if you take the expectation of the random vector x, here x is the random vector. And why? Because x comprises all these individual random variables. So, x1, x2 up to xp. So, if you want to find out the mean of the entire random vector, we can actually find out the individual means.

So, E of x1, E of x2, E of x3, so on and so forth up to the expectation of xp, and then the transpose of that. So, this essentially becomes something like mu1, mu2, mu3, mu4 up to mup. Okay. Now, again, just one point to note here is that this denotes a transpose. And why transpose?

Because generally, we consider vectors to be column vectors, right? So, if you want to write down the vectors as row vectors, then you put a transpose to ensure that the vector is indeed a column vector. So, if you transpose a row vector, it becomes a column vector. Make sense so far? Okay.

And then the second thing is, what would this be? So, let us say x minus mu, and then x minus mu transpose, and then the expectation of that. So, expectation of x minus mu into x minus mu transpose. By the way, what is mu here? So, mu is nothing but this entire thing—a random vector.

So, this could be denoted by the vector mu, right? So, what did we start with? So, we started by finding out the expectation of the random vector x, which is nothing but the individual expectations of each and every random variable inside that random vector, which could be written down as individual means. So, mu 1, mu 2 up to mu p, but this, in fact, is another random vector, which could be denoted by mu. And now what we are doing is we are finding out the covariance in a way, isn't it?

Because expectation of x minus mu into x minus mu transpose. So, and then this should in fact be a matrix, isn't it? Because x minus mu is a random vector, x minus mu transpose is another random vector, right? So, a random vector multiplied by its transpose has to be a matrix, right? And then this is exactly the matrix that it should be.

So the first entry, the 1, 1 entry should be expectation of the first random variable minus its mean whole square. So essentially what is this? This is nothing but the variance of x1, isn't it? So this is the variance of x1. Similarly, the second diagonal entry would be variance of x2.

And similarly, the last diagonal entry, this is nothing but variance of the random variable xp. So, in a way all the diagonal entry should be the corresponding variances and all the off diagonal entry should be the corresponding covariances. For example, this. So, what exactly is this? This expression is nothing but the covariance between x1 and xp.

Isn't it? As per the formulas. So, this is nothing but the covariance between x1 and xp. Similarly, this thing is nothing but the covariance between xp and x1, which is exactly equal to that, by the way. So, similarly, all the off-diagonal entries would be the covariances, and all the diagonal entries would be the variances.

So, this is the famous matrix that we have. So, the diagonal entries are sigma 11, sigma 22, sigma 33, up to sigma pp. And all the off-diagonal entries are sigma 1, 2, sigma 1, 3, up to sigma 1, P, and similarly others. So, what should be the order of this matrix? The order of this matrix is P cross P—so, P rows and P columns.

$$E(X) = \left[E(X_1), E(X_2), \dots, E\left(X_p\right)\right]' = (\mu_1, \mu_2, \dots, \mu_p)'$$

$$E(X-\mu)(X-\mu)' =$$

$$\begin{bmatrix} E(X_1-\mu_1)^2 & \cdots & E(X_1-\mu_1)(X_p-\mu_p) \\ \vdots & \ddots & \vdots \\ E(X_p-\mu_p)(X_1-\mu_1) & \cdots & E(X_p-\mu_p)^2 \end{bmatrix} =$$

$$\begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{bmatrix}$$

And in other terms, such a matrix is also called a dispersion matrix. So, this is also called a variance-covariance matrix, by the way. So, this matrix is called a variance-covariance matrix. Because it sort of combines all the variances and all the covariances among the random variables inside a random vector. And the other name for a variance-covariance matrix, in time series terminology, is also a dispersion matrix.

So, hopefully, all these elementary ideas are clear before we proceed. Now, what do you think a cross-covariance matrix means? So, cross-covariance means that if you have two random vectors, let us say x and y. So, earlier, whatever we have discussed so far was only for a single random vector. Now, imagine you have two different time-series processes.

So, let us say one is xt and the other one is yt. Now, as we discussed earlier, each one of these is a random vector. So, xt is a random vector, and yt is another random vector. So, this, in fact, falls in the framework that we are studying in this slide. So, let us say if you have two random vectors x and y with corresponding mean vectors mu x and mu y, then the cross-covariance matrix is defined by nothing but the covariances between the two random vectors.

So, the expectation of x minus its mean vector multiplied by y minus its mean vector and then transposed. If you remember earlier, we had x minus mu multiplied by x minus mu transposed. So, both the random vectors were kind of the same. But here, if you have two random vectors, we want to try to find out the interrelationships or the cross-covariance between those two. So, this is a time-series terminology again, which is called cross-covariance.

So, cross-covariance is the covariance between two different time series or two different random vectors. And hence, now we will study some properties of this cross-covariance matrix and so on and so forth. And then, by the way, this is a notation that we stick to. So, sigma xy, sigma subscript xy. So, sigma subscript xy denotes the cross-covariance matrix of two random vectors x and y.

So, the ijth element of this matrix becomes nothing but this. So, if you take the ijth element, the ijth cell of that matrix, that is nothing but the expectation of xi minus mu xi into yi minus mu yj transpose, which is nothing but the covariance between these two random variables. So, xi from the first random vector and yj from the second random vector. So, such an expression or the ijth element of that cross-covariance matrix is nothing but a single covariance between two random variables. So, xi and yj coming from these two individual random vectors.

$$\Sigma_{xy} = E(X - \mu_x)(Y - \mu_y)'$$

Now again, I am very sure that all these ideas would be slightly new to all of you if you do not have a background in statistics or something like that. But then you want to revise them multiple times. So, the idea of random vectors, the basic idea of linear algebra. Now, how do you transition from a single random vector to multiple random vectors and so on and so forth? Okay, so now how do you transition all these notations to a time series sort of framework?

So, we define what you call a stationary multivariate time series initially. Okay, so let us say yt is a time series, and then yt equals y1t, y2t up to ykt transpose. Okay. And by the way, each one of these that you have here is nothing but an individual time series process. So, T is a timestamp.

So, T belongs to capital T, which could be either plus or minus 1, plus or minus 2, plus or minus 3, etc., and so on. So, essentially, this yt denotes a k-dimensional time series vector at time t. Again, just to repeat, this yt is a collection of k time series processes. So, k individual time series processes. So, y1t, y2t, y3t up to ykt. So, for that matter, each one of these is nothing but an individual time series. So, y1t, y2t up to ykt.

And hence, this is a k-dimensional time series vector at time t. So, when can we say that this collection of k time series is stationary? So, the process yt is stationary if the joint probability distribution of yt1, yt2 up to ytn is the same as the joint probability distribution of yt2. The same series expressed with some lag. So, let us say yt1 plus l, yt2 plus l, yt3 plus l up to ytn plus l. So, here, if you notice, the only thing we are doing is shifting each timestamp by some lag, which is l in general. So, if the underlying distribution of this first collection happens to be exactly the same or exactly similar to the underlying distribution of the bottom collection, where we are shifting each timestamp by some lag L, then we can say that this original collection, which is Yt, is stationary.

So, the idea, the main idea here is that we are not expressing stationarity of a single time series, because Yt is not a single time series here. What is Yt? So, Yt is a collection of k different time series, isn't it? We have shown that earlier. So, Yt is a collection of k different time series: Y1t, Y2t up to Ykt.

And now, how can we say if that collection itself is stationary? So, this is a slightly different idea from what we have studied earlier. And obviously, this should be true for all timestamps. So, T1, T2 up to Tn and all the leads or lags which are L—so, 0 or plus and minus 1, plus and minus 2, etc. So, my suggestion would be that you should actually proceed ahead if you have understood very clearly the idea of stationarity for a multivariate time series.

Make sense now? So, now can we put forward some expressions for moments of the process? So, assume that the process is stationary. So, assume that Yt is stationary. Now, again remember Yt is not a single process.

So, yt is a collection of several time series or, for that matter, k different time series. So, assume that process is stationary, and the mean vector and the covariance matrix are finite. So, the mean vector of yt and the covariance matrix of yt are finite. So, these are just the notations. So, the mean vector of the process could be denoted by the expectation of yt, which is nothing but mu, which is nothing but a collection of individual means.

So, mu1, mu2, mu3, mu4 up to muk. Now, again, just to summarize, what are each of these individual values? So, for example, what is mu1? So, mu1 is nothing but the mean of the first time series inside that collection, right? Mu2 is nothing but the mean of the second time series inside that collection.

And similarly, muk is nothing but the mean of the kth series inside that collection. So, if you want to write down the expectation of the entire random vector yt, that would be nothing but this collection of individual means. And similarly, we can talk about, let us say, the covariance matrix. So, which is nothing but the expectation of yt minus mu multiplied by yt minus mu transpose, and then, notation-wise, this would be sigma. So, hence, we can actually define some moments of the process, which is yt.

And then immediately we can talk about cross-covariance or the cross-correlation. So, again, assume that the process is stationary. So, for a stationary process, the covariance between y_i(t) and y_j(t + l) depends only on the lag l and not on time t. So, I think this property we studied long back, is it not? I mean, when do you say that a process is stationary by looking at the covariance? If the covariance only depends on the lag and not the actual timestamp t, right?

So, again, if you look at these two entities, the lag is l, right? So, y_i(t) and then y_j(t + l), but then the lag is l. When it comes to timestamps, so t and then t + l, right? So, if the covariance between these two depends only on l and not on t, and of course, this should be true for all i, j, all l, t, etc., then we can say that the process is stationary, all right? Now, we can actually define these two entities.

So, the auto-covariance function of y_i(t), right. So, what should be the auto-covariance function of y_i(t)? Now, the notation should be gamma_ii at the lag l. So, here, remember one thing: when you are proposing or defining any covariance or correlation, there should always be some lag. So, we are trying to sort of put forward a formula for y_i(t) and then y_j.

And then let us say T plus L. So, you have some lag between the two values. So, the auto-covariance function of YIT would be denoted by gamma II of L. And then this is nothing but one can actually apply the standard formula. So, the expectation of YIT minus its mean into yi t plus l minus its mean. Now, by the way, one thing to note here is that the auto-covariance function of yi t is gamma ii l, and then yi t here is a single process, a single time series. So, among all the k different time series we saw earlier.

So, I am sort of helping all of you to understand the notation. So again, remember where did we start? So, we started with defining yt. And what exactly is yt? So, yt is a collection of all these individual time series processes.

So, y1t, y2t up to ykt. Make sense? So, how many time series processes do you have in this collection? You have k, basically. And yt is a combination of all these individual time series.

So, y1t, y2t up to ykt. Now, initially here, what we are trying to put forward is a formula for the autocovariance function of a single time series from this collection. So, let us say yit. Okay. Yit is what?

So, yit is one of these. So, this is at the ith position, by the way. Right. So, can we define the autocovariance function of a single time series out of this entire collection? Right.

And for that, we have this notation. So, gamma iil, which is nothing but the expectation of yit minus its mean. So, its mean should be mui because we are talking about yit, right? And then, yi comma t plus l minus its mean. Now, again, its mean is mui because both the indices are the same, right? Because here, the focus is only on yit, right? Now, the only thing we are doing is we are basically shifting the timestamps by some lag. So, yit and then yit plus l, and then we want to find out the covariance between these two. Hopefully, you understood the idea that initially, autocovariance is always defined on a single series.

But now, the second idea which we want to talk about, which is different, is: can we define cross-covariance? So, cross-covariance between any two series. So, YIT comma YJT at some lag. So, for example, I can talk about the cross-covariance between Y1T and Y2T. Right? Isn't it?

Or cross-covariance between Y1T and Y3T. So, take any two series from this collection, let us say Y1T and Y2T, and then we can find out the cross-covariance between these two-time series at a particular lag. So, what exactly is the definition for that? So, the notation would be gamma IJ L because now both these indices will not be I, right? Because we have YIT and then YJT as two individual time series. So, gamma IJ at the lag L is nothing but the expectation of YIT minus its mean. So, its mean would be mu I because the index is I, and then YJ comma T plus L minus its mean, and then its mean should be mu J now because now the index is changing.

Make sense so far? So again, my task is to sort of simplify all the notations for you. So again, if your focus is on a single series, then it should only be focused on the single index. So, either I or J, whatever. But the moment you bring in more than one series.

So, let us say two series; then, the cross-covariance should depend on two indices. So, i and then g. Similarly, one can talk about the cross-correlation. So, how do you define the cross-correlation function? So, the cross-correlation between y_i(t) and y_j(t) again at lag L is nothing but the cross-covariance at the top divided by the individual variances or the individual autocovariances, which are nothing but gamma_ii at lag 0, gamma_jg at lag 0, and then the square root of that. So, now if you notice one thing, all the formulas are pretty standard.

So, how do you find out covariance? Expectation of something minus its mean into the difference series minus its mean, right? Or how do you find out correlation? So, again, this is a standard formula. So, correlation is autocovariance divided by the square root of the corresponding variances, right? Similarly, since we want to define the cross-correlation, cross-correlation is always between two series, right?

So, Y_i(t) and Y_j(t). So, the numerator would be the cross-covariance. So, the numerator here is not the autocovariance of a single series, but this is the cross-covariance, and then divided by the individual autocovariances of Y_i(t) and then Y_j(p). But now the idea is that we can actually get this cross-covariance matrix, right, where again the same idea. So, the diagonal entries are what?

The diagonal entries are the individual autocovariance functions, right. So, gamma 11L, gamma KKL, etc. And then all the off-diagonal elements are nothing but the cross-covariance is among two series. And similarly, we can get the cross-correlation matrix also. So, cross-covariance matrix and then the cross-correlation matrix.

So, how would the cross-correlation matrix look like? So, again definition is rho L and here we can actually standardize the formula by bringing in some diagonal matrix. So, let us say V capital V and then V to the power minus half into the cross-covariance matrix gamma L into V to the power minus half where V is a diagonal matrix comprising of all these entities. And these are some interesting properties that one has. So, rho i i l equals rho i i minus l. So, if you change the lag from l to minus l. So, something like let us say t plus 2 and then t minus 2.

So, correlation between y t and y t plus 2, should be the same as yt and yt minus 2. So, this is the meaning of that or rho ijl should be equal to rho ji minus l or gamma ijl should be equal to gamma ji of minus l. So, these are some interesting properties about the cross-covariance matrix or the cross-correlation values etc., And hence, we can actually have these properties. So, gamma L equals gamma of minus L transpose and similarly rho L equals rho minus L transpose. The cross-covariances and cross-correlation hence develop the dynamic relationships between any two times.

So if you want to analyze two different time series processes, or rather more than one time series process, then as we studied earlier in the last lecture, we have to transition from analyzing a univariate series to either bivariate, trivariate, or multivariate. And there, the idea of a single autocorrelation does not serve the purpose, right? Because now we have more than one. So, there the idea of cross-covariance or cross-correlation comes in handy. Make sense?

So, this is the main idea to study the interdependencies or the interrelationships between any two individual time series processes. Now, the second definition we will focus on in this session is weak second-order stationarity. So, when can we say that a time series process is second-order stationary? So, a process yt is second-order stationary if the expectation of yt is mu. So, you do not have a subscript here.

So, this does not depend on t, of course. And if you talk about the cross-covariance matrix of gamma l, which is nothing but the expectation of yt minus mu and then yt plus L minus mu transpose, it only depends on the lag L, okay? Now, by the way, another name for such stationarity is covariance stationarity, right? So, if you want, you can write it down. So, this is also called covariance stationarity, okay?

So, of course, two points to remember here. So, the mean should not depend on t, and if you talk about the cross-covariance matrix, that should only depend on the lag, which is L, all right? Okay, so now that we have defined some properties about random vectors, covariances, correlations, etc. Now, we can actually define the very first multivariate process, which is the vector white noise process. Now, again, we will start with some basic processes, such as white noise, and then try to build upon these processes.

So, what do you mean by a vector white noise process? So, let us say E_t. So, E_t is a collection. E_t is defined as WN(0, sigma). So, WN stands for white noise, and the mean vector is a 0 vector, and this is nothing but the variance-covariance matrix. So, we can say that the collection E_t follows a vector white noise process if and only if E_t is stationary with a mean 0 vector and the variance-covariance matrix given by this structure.

So, gamma_k equals sigma whenever k is 0, and for any other value of k, the gamma_k value should be 0. This is kind of similar to how we defined a white noise process earlier. If you remember, a single white noise process, let us say E_t, because this is the general assumption that we make. So, E_t is a random error. So, we can say that E_t is white noise with mean 0 and, let us say, variance sigma squared E. So, we have been defining it this way until this point.

But if you want to transition to, let us say, more dimensions, bivariate, trivariate, multivariate, etc., So naturally you have to bring in the idea of a mean vector and a variance covariance matrix. So again just to summarize this that this is a collection of several errors now. So ET is a random vector. So, the random vector follows a vector white noise process.

So, W n 0 comma sigma if and only if E t is stationary with mean 0 vector. So, mean should be the 0 vector and variance covariance matrix given by this structure. So, gamma k equals sigma whenever k equals 0 and 0 otherwise. So this way probably we will try to extend this to many other vector models. So soon probably in the next sessions which are coming in this week we will talk about let us say vector AR models or vector MA models.

So in short we call them as VAR or VMA. And if you join these two they become vector ARMA models. So VARMA. So in a way all these are extensions of the individual AR models or MA models or ARMA models and so on ok and of course again like I said do not worry too much that towards the end we will sort of tie all these things down by working through a practical data in R also and probably there you will try to will try to sort of understand all these ideas or all these individual ideas in a slightly better manner. So as you will get the connection between what is going on in theory here and then the actual practical aspect of that.

Thank you.