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 33: Some Specific Multivariate Time Series Models

Hello all, welcome to this course on time series modeling and forecasting using R. Now again, we are in the middle of this week, and the focus area we are discussing this week is all about multivariate time series analysis, right? So, if you vaguely remember, in the last couple of sessions, we sort of introduced the idea of why we want to transition from a univariate time series process to something called, let us say, either bivariate or multivariate time series analysis, right? And then, even the last lecture— If you do not remember, then what exactly do you mean by a multivariate time series process? Probably, I will very quickly revise that, and then we will sort of move on from there. Now again, the whole idea of explaining multivariate time series analysis to somebody is not really easy, right?

So, my strong suggestion is again that you should go back and try seeing some of the initial videos in this week and then sort of repeat viewing those sessions, right? Just to sort of understand that if you are transitioning from a univariate Yt to a notation which is multivariate Yt, then what differences do you have to take care of, okay? So, for example, what do I mean by that statement is that initially, up to now, we have discussed that let us say Yt is a time series process, and then one particular realization of Yt could be something like, let us say, Y1, Y2 up to Yn, right? So, let us say you take an example of any univariate time series example. So, let us say temperature data or stock price for a single stock, and so on and so forth, right?

$$Y_t = (Y_{1t}, \dots, Y_{kt})' : t \in T \equiv (\pm 1, \pm 2, \dots)$$

So, let us say Yt denotes the temperatures, right? Let us say Yt denotes the monthly temperatures for a year. So, all these would be the monthly temperature values which are realized, or Y1, Y2 up to Yn is nothing but the collected data. So, this is an example of a univariate time series. But the moment you transition from univariate to multivariate, now let us say the notation changes very slightly, something like Yt inside curly brackets.

And then now yt becomes a collection of random vectors. Even here yt, when you talk about a univariate situation, yt is a single random vector. But now when you talk about yt, it is a collection of several random vectors. So, let us say I can write something like y1 and then y2 up to y, let us say k. So, my yt, which is a multivariate time series process now, is a collection of a k-dimensional random vector. So, y1, y2, yk in themselves are nothing but individual time series or individual univariate time series.

And again, properties of any multivariate time series process like yt, then interconnections between any two vectors here, so let us say y1, y2, right? So, all these properties have to be carefully defined. And hence, we have a notion called, let us say, cross-covariance. So, in the last lecture, if you remember, we discussed an idea about cross-covariance. So, cross-covariance means nothing but the interrelationship between any two series of the same collection of random vectors, which is yt.

Okay, so now the slide that you see in front of you discusses some results on a technique called linear filtering. Now we will quickly try to understand what you mean by, let us say, filtering, and then what you mean by linear filtering, and what all are the ingredients that go inside a filtering. Okay. So, now let us say that you have xt, which is an r-dimensional input series. So, xt denotes the input series, and let yt be a k-dimensional output series.

$$Y_t = (Y_{1t}, \ldots, Y_{kt})' : t \in T \equiv (\pm 1, \pm 2, \ldots)$$

So, assume that you have some input variables, and then you have some time series model or time series structure, which outputs a series. So, xt is a collection of r such random vectors, which is the input series, and similarly, yt is a collection of k such random vectors, which is called the output series. Then, how can we define a multivariate linear, also known as a time-invariant filter, relating xt and yt? So, in other words, how do we define some relationship between this input series xt and the output series yt? And this is exactly that equation.

$$Y_t = \sum_{j=-\infty}^{\infty} \Psi_j X_{t-j},$$

So, let us say yt equals some infinite sum, j going from minus infinity to infinity, and then a collection of some coefficients, let us say ψj, applied on xt minus j. So, again, just to reiterate, xt is the input series, yt is the output series, and through this equation, we are creating that multivariate linear filter. So, in other words, such an expression is called a filter, and 'filter' is a technical term in time series, which denotes the interdependency between an input series and an output series. And again, since we are dealing with a multivariate structure, all these ψj's are themselves matrices. So, let us say k × r matrices, and why k × r? Because both these input and output series are r-dimensional and k-dimensional, respectively.

So, this is the definition, or this is the expression, of a particular multivariate linear filter. Now, again, what would happen if ψj is 0 for all indices which are negative or for all j less than 0? Now, again, just for a second, if you go back to the slide we saw earlier. Now, again, remember that the summation is running from minus infinity to infinity. And now, on the next slide, we are assuming that all the negative indices.

$$\text{If } \Psi_j = 0 \ \forall \ j < 0,$$

$$Y_t = \sum_{j=0}^{\infty} \Psi_j X_{t-j}$$

So, in other words, if j is less than 0, then the corresponding value of psi j is 0. So, in that case, what would happen is that the summation would be narrowed down to only positive terms. So, j would be running from 0 to infinity, but the structure remains the same. So, y t equals the summation of j going from 0 to infinity psi j and then x t minus j. In this situation, the filter is physically realizable or causal. So, again, these are some technical terms we put forward for the filter.

So, in this case, we call the filter physically realizable or causal. And in this case, y t is expressible only in terms of present or past values of the input series x t. And naturally so, right? So, if the summation is only running from 0 to infinity, let us say what would happen if you put 0? So, if the value of j is 0, this becomes x t, right?

Then, if you put the value of j to be 1, this becomes x t minus 1, then x t minus 2, and so on, isn't it? So, essentially, the values which y t depends on are nothing but either the current value of x t or its past values, but not the future values. So, this is the statement that we have here. So, in this case, y t is expressible only in terms of present and past values of the input series x t. Now, another definition we will put forward is when we can say that a linear filter is stable. So, you have an understanding or a notion of a filter being stable.

So, for that, we require another ingredient called a norm. So, let this notation denote the norm of any matrix A, which is defined as nothing but the norm of A is nothing but the trace of A transpose A. Now, here, assume that A is a matrix, right? So, this is the notation for the norm of A. And then the actual expression is the norm of A is nothing but the trace of A transpose multiplied by A. So, under this condition, if the summation j going from 0 to infinity of the norm of psi j is finite, then we say that the underlying linear filter, which is given by this expression, is stable, right? So, this is a small definition of when we say that a linear filter is stable—whenever this condition is satisfied.

$$E(Y_t - \sum_{j=-n}^{n} \Psi_j X_{t-j})(Y_t - \sum_{j=-n}^{n} \Psi_j X_{t-j})' \to 0, \text{ as } n \to \infty$$

Now again, I am pretty sure that all the notations are slightly more difficult compared to earlier sessions. Let us say when we deal with only univariate situations, which is natural, so. So, if you are transitioning from univariate to, let us say, even bivariate or multivariate, then you have to have some understanding about the dependencies between, let us say, matrix multiplications or the idea of inverse, adjoint, which we will soon see, of course, right? But then you have to sort of have some basic idea about matrix theory or linear algebra, okay? So, now, if the linear filter is stable and the input random vectors xt have finite second moments.

So, if the linear filter is stable—in the last slide, we saw when we say that a linear filter is stable—and on top of that, if the input random vectors xt have finite second moments, then for the output vector yt, the filtering representation exists uniquely and converges in its mean square. What do you mean by that? So, converging in mean square means if you take the expectation of yt minus whatever you have on the right-hand side and then the same thing and the transpose of that. So, this should approach 0 as n tends to infinity, right? So, under this condition, which is given by star, we will say that the filtering representation exists uniquely and it also converges in mean square, okay?

And now, on top of that, we can define, let us say, the covariance matrices or the correlation matrices, and so on and so forth. So, if the linear filter is stable and Xt is stationary with cross-covariances denoted by, let us say, gamma XL, then it also means that Yt is also stationary with cross-covariances given by this formula. So, gamma YL is nothing but the covariance between Yt and Yt plus L. Now, again, there is nothing new here because this is just the definition of cross-covariance, and then the expression for that is given by this structure. So, here you have two summations running from minus infinity to infinity, and then the psi i coefficients, then the gamma variable in terms of x. So, this expression that you see in between is nothing but the cross-covariance for the random vector xt, isn't it? So, the subscript x denotes anything related to xt, and the subscript y denotes anything related to yt, okay?

$$\Gamma_y(l) = COV(Y_t Y_{t+l}) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \Psi_i \Gamma_x(l+i-j) \Psi_j{}'$$

So, this is a not-that-easy structure, I will say, but then nevertheless, it sort of defines the cross-covariances among the random vectors comprising yt, ok. So, these are just some results on linear filtering. So, let us say, for example, when can we say that the filter exists uniquely, or when do you say that the filter sort of converges in mean square, right? And further, how do you find out, let us say, the covariance matrix or the cross-covariances, and so on and so forth, ok. Alright, so now we will slowly transition to a slightly different trajectory by defining some models.

So, now that we have some understanding about what you mean by, let us say, multivariate time series structure, and so on and so forth. So, now is the correct time to sort of define or put forward some multivariate time series processes. So, the first representation is given by Wold, and hence the name is Wold or MA representation of the series. And here again, it will be a good idea if you go back to the MA session from the univariate structure and then just sort of understand how an MA process looks like. Now, again, just to give you a very quick refresher, an MA process is nothing but you have the time series yt on one side and the moving average in terms of all the random errors on the right-hand side.

So, similarly here, something of a similar nature should happen. So, even if you transition from a univariate MA to, let us say, a multivariate MA, things will not change very drastically. So, what exactly do you mean by this Wold representation? This is, in fact, also called an infinite MA representation of a stationary vector process, and why infinite, we will soon see. Now, again, let us assume that yt is a multivariate stationary process.

So, again, yt, as before, is a collection of some random vectors with, let us say, a mean vector mu. So, yt is a k-dimensional collection of random vectors. So, yt is k cross 1, and similarly, the mean vector has to be k cross 1. Then this process, so yt minus its mean vector, equals an infinite sum—j going from 0 to infinity—psi j e t minus j, is given by nothing but the Wold representation or the infinite MA representation of the series. Now, a couple of things here: firstly, why infinite? So, infinite because you have an infinite sum.

$$Y_t - \mu = \sum_{j=0}^{\infty} \Psi_j e_{t-j}$$

$$Y_t = \mu + \sum_{j=0}^{\infty} \Psi_j e_{t-j} = \mu + \Psi(B)e_t$$

And the second question is, why MA? So, why MA? Because if you take this mu on the right-hand side, then this is nothing but the MA structure. So, you have yt on one side and nothing but the moving average or collection of some random errors on the other side, which is nothing but the MA representation. Make sense?

So, this is exactly what we will do now. So, we will sort of transport the mu on the right-hand side, and then yt is nothing but mu plus the infinite sum, where, of course, what is psi b? So, psi b is nothing but the summation psi j b to the power j. So, this psi b is nothing but a k cross k matrix of backward shift operators. So, again, remember this capital B denotes the backward shift operator. So, again, remember this capital B denotes the backward shift operator, right?

And then, if you are representing in a multivariate sort of structure, then you should have a matrix. So, a K cross K matrix of backward shift operators, okay? Alright, so now the next module we will discuss is a more generalized version, which is called the vector autoregressive moving average. Or, in short, we will call that a VARMA model. So, VARMA.

So, again, VARMA should kind of have a similar structure to a univariate ARMA process, isn't it? The only difference is there will be some matrices there, there will be some random vectors there, and so on and so forth. So, in general, a VARMA P Q process could be denoted by this expression. So, you have phi P B on the left-hand side applied on yt minus mu equal to theta Q B applied on the random errors, which is et.

Now, remember one thing: whenever you see et, all these ets are not univariate, right? Because we have a multivariate structure.

So, all these ETs are assumed to be white noise multivariate processes. So, white noise, or rather multivariate white noise processes. In the last session, we sort of discussed what you mean by a multivariate white noise structure, OK. So, all these random errors are no longer univariate, but they take a multivariate sort of structure, OK. But again, all the other things, if you notice in this slide, are sort of similar to when we discussed ARMA processes in a univariate case, isn't it?

Because again, this phi p b coefficient is nothing but phi 0 minus phi 1 b minus so on and then phi p b to the power p. And similarly, theta q b coefficient is nothing but the same thing. So, theta 0 minus theta 1 b minus so on and so forth up to theta p and then b to the power p. Now, here the only difference is all these coefficients, right? So, phi 0, phi 1, phi p or theta 0, theta 1, theta p. So, these are some matrices, OK? And then these are not some single constants.

$$\Phi_p(B)(Y_t - \mu) = \Theta_q(B)e_t,$$

$$\text{Where, } \Phi_p(B) = \Phi_0 - \Phi_1 B - \cdots - \Phi_p B^p$$

$$\text{And, } \Theta_q(B) = \Theta_0 - \Theta_1 B - \cdots - \Theta_p B^p$$

$$q = 0 \Rightarrow \Phi_p(B)(Y_t - \mu) = e_t \Rightarrow VAR(p)$$

$$p = 0 \Rightarrow (Y_t - \mu) = \Theta_q(B)e_t \Rightarrow VMA(q)$$

So, these are matrices now, OK? Alright, so some specialized cases, so what would happen if you plug in q to be 0? So, if q is 0, then the equation reduces to this form. So, phi p b applied on y t minus mu happens to be only the multivariate white noise structure, which is e t. And this model is nothing but VAR with an order p, right?

So, you do not have any m a structure or you do not have any m a component here. Similarly, on the other hand, what would happen if P happens to be 0? So, when P is 0, this implies that Yt minus mu equal to theta Qb coefficient applied on Et and this process is nothing but Vma with an order of Q. Okay, so when we put forward an equation for Varma PQ process, which is general looking vector autoregressive moving average, then depending on the fact that whether my P could be 0 or Q could be 0, I can sort of reduce the models to be either VARP or VMAQ. Okay, so now further we will try to understand some properties about this V ARMA process.

So, firstly this process would be stationary if it can be represented as a convergent vector moving average process of infinite order. Now, firstly what do you mean by that? So, again this is the structure written in a slightly different way. So, let us say y t equals mu plus e t plus summation j going from 1 to infinity, and then some psi j coefficients applied on E t minus j, right? Now, again, just for a second, if you compare this equation to the equation that we had earlier, right?

$$Y_t = \mu + e_t + \sum_{j=1}^{\infty} \Psi_j e_{t-j}$$

Or, $Y_t = \mu + \Psi(B)e_t$, where $\Psi(B) = 1 + \sum_{j=1}^{\infty} \Psi_j B^j = \Phi(B)^{-1}\Theta(B)$

So, the equation for the V-ARMA process is given by this, where you have a phi p b coefficient and then y t minus mu and then theta q b coefficient applied on E t, okay? Now, just for a second, let me write down the same equation here and then we will try to match them. So, phi p b and then y t minus mu equals let us say some theta q b coefficient applied on it. So, this is the exact equation from the last slide and then we want to sort of compare this equation and then this equation. Isn't it?

Because this is just another way of writing down the same VARMA equation. So here, if you notice, what we are doing is we are trying—so first thing is we are bringing this mu to the other side, right? The other thing is we are taking an inverse of this phi pb coefficient and then moving them to the RHS. So what would happen if you take this phi pb coefficient to the RHS? You have to take an inverse of that, okay?

So essentially, yt is nothing but mu plus et, right? Plus again this infinite sum j going from 1 to infinity, and then psi j coefficients applied on E t minus g. Now, the whole thing is: what is psi j, right? So, the whole thing is: what is psi j? And then, from discussing what psi j is, we will have some connection between these two equations, okay? Or in other words, we can write down y t to be nothing but mu plus psi d coefficients applied on E t, where the psi b coefficients are nothing but 1 plus the same summation j going from 1 to infinity psi j b to the power j, and then this expression is nothing but my phi b coefficient and then inverse of that into theta b coefficient. Now, again, what exactly is happening here? I will maybe explain in a minute or so. So, from this equation, can you see that y t minus mu can be written down as phi P B coefficient inverse, is it?

And then theta B coefficient applied on E T. Does that make sense? So, I am simply taking the inverse of the matrix phi P B and then moving that to the right-hand side. And now, from this equation, I can move mu to the other side. So, y t happens to be mu plus my phi P B coefficient inverse of that and then theta Q B coefficient applied on E T. Does that make sense so far?

So, this equation is essentially nothing but this expression that we are trying to simplify in the last line, isn't it? Because here, y t is what? y t is mu plus psi b e t. So, again, if you compare this equation and the first equation you have here, right? So, essentially, psi b should be nothing but equal to phi P B inverse into theta cube B, isn't it? Because this should match, right?

So, phi P B inverse theta cube B should be nothing but exactly equal to my psi B coefficient, which is nothing but what we are trying to explain here. So, psi B eventually is nothing but equal to my phi B inverse into theta B. Now, the whole point of rewriting the initial equation into a different form is that the most important point here is that such a structure is stationary. So, the process is stationary, as stated here also, that the process is stationary if it can be represented as a convergent vector moving average process of an infinite order. Now, what extra thing is special about this structure is nothing but this is an infinite MA process.

Can you see that? Because on the right-hand side, you do not have any YTs, right? So, this is nothing but a combination of simply errors, isn't it? So, this is nothing but a VMA process of an infinite order. So, an infinite moving average process.

So, essentially, just to summarize this slide: when can we say that a VARMA process is stationary? If a VARMA structure can be represented or rewritten as a convergent vector moving average process, or in other words, a VMA process of infinite order. Okay. So, I guess this is just a rule for determining whether a VARMA process is stationary or not, etc. Okay.

Okay, so I think this slide is just a small refresher about, let us say, what do you mean by determinant or what do you mean by adjoint, right? So again, let us say A is a matrix whose determinant is denoted by the determinant of A and the adjoint by the adjoint of A. Then, of course, you must have studied back in, let us say, early college days when you took a course on linear algebra that the inverse of the matrix can be written as something like 1 divided by the determinant multiplied by the adjoint. Now, again, specifically, I am suggesting that if you do not have any background in linear algebra or matrix theory, then probably pick up any basic book, and all these identities would be there. But using this relation, we can actually write down a small expression for the inverse of this phi z matrix.

And why are we more concerned about the inverse of the phi z matrix? Because again, if you go back for a second, we have exactly this matrix with us, don't we? So, it would be a nice idea to express this inverse matrix using, let us say, determinant or adjoint and so on. So, I can write down the inverse of this phi z matrix as nothing but 1 divided by the determinant. So, I have the determinant inverse multiplied by phi star z, where phi star z is nothing but the adjoint.

$$A^{-1} = \frac{1}{\det(A)} adj(A).$$

Thus we can write, $\Phi(z)^{-1} = \det(\Phi(z))^{-1}\Phi^*(z)$, $\Phi^*(z) = adj(\Phi(z))$.

The process can be written as, $Y_t = \mu + det(\Phi(B))^{-1}\Phi^*(B)\Theta_q(B)e_t$.

The process is stationary if $\{\det(\Psi(z))^{-1}\}$ is convergent for $|z| < 1$.

Thus, if you go back to the earlier slide again, I can actually write down the V-ARMA equation in this form. So, y_t is μ plus the determinant of φ(B)^-1 multiplied by the adjoint. So, φ*(B) multiplied by the θ_q(B) coefficient applied on e_t. Now, the only thing I am doing here is expressing the expression on the last slide in the form of 1/determinant multiplied by its adjoint. And again, the whole idea is that the process is stationary if this term, the determinant of ψ(z)^-1, is convergent whenever the absolute value of z is less than 1.

So, you have one more simpler-looking idea as to when you can say the process is stationary or not. So, when you write down such an expression containing determinants or adjoints and so on, the only thing you have to prove is whether your 1/determinant or determinant^-1 is convergent for every value such that |z| < 1. Alright, so once you define a general VARMA process, we can then pinpoint some specialized processes. So, the first one is, let us say, VMA of order Q. This process is nothing but VMA of order Q or a vector moving average process of order Q. Okay.

So, here we can fix. So, rather than fixing the upper limit to be infinity, we can fix it to be Q. Okay. And then again, we will go back to the earlier equation. And then, how can you write down that equation? So, y_t would be nothing but μ plus e_t plus the summation from j=1 to Q of θ_j multiplied by e_{t-j}. Or, I can write this in a concise form.

$$Y_t = \mu + e_t + \sum_{j=1}^{q} \Theta_j e_{t-j} = \mu + \Theta(B) e_t$$

$$\Theta(B) = I_k + \Theta_1 B + \Theta_2 B^2 + \cdots + \Theta_q B^q$$

So, mu plus theta b coefficients applied on E t, where the theta b coefficients are nothing but given by that expression. So, you have an identity matrix of order k plus theta 1 b plus theta 2 b square and so on and so forth up to theta q b to the power q. So, this theta b coefficient that you see here is nothing but given by that expression. So, such a structure is called a VMA of order q or a vector moving average process of order q. So, similarly, we can talk about some specialized examples. So, let us say a vector MA1 process.

So, this is VMA of order 1, right? Now, again, just for simplicity, we can assume that the mean is 0. So, again, we can invoke some recursive substitution here. So, this is the equation for VMA1, is it not? So, again, if you are not very comfortable, you can pause the video and then just check.

$$Y_t = e_t + \Theta_1 e_{t-1}$$

$$= e_t + \Theta_1 (Y_{t-1} - \Theta_1 e_{t-2}) = e_t + \Theta_1 Y_{t-1} - \Theta_1^2 e_{t-2}$$

$$= \cdots$$

$$\Theta_1 Y_{t-1} + \Theta_1^2 Y_{t-2} + \cdots + e_t \equiv VAR(\infty)$$

So, this is my VMA1. So, y t equals e t plus theta 1 e t minus 1, okay. Now, here, what I can do is I can express or re-express every time this E t minus 1 as Y t minus. So, from here, what can I do? So, my E t minus 1 is nothing but Y t minus E t and then divided by theta 1, is it not?

So, what would happen if I go on sort of putting this expression every time here? So, the equation becomes E t plus theta 1 Y t minus 1 minus theta 1 E t minus 1. And why is that? Because from this equation, right, from that equation, can you see that I can write down et as what? So, from this equation, et becomes yt minus theta 1 et minus 1.

Isn't it? And then, E t minus 1 would be nothing but y t minus 1 minus theta 1 and then E t minus 2. And then this I can replace back where I have E t minus 1. And here, I can sort of have a recursive substitution.

So, the first stage would be that right, and again I can replace e t minus 2 similarly, and so on and so forth, ok. So, eventually the expression would be nothing but that, which is theta 1 y t minus 1 plus theta 1 square y t minus 2, and so on and so forth up to e t, and this expression, if you see closely, is nothing but the expression for a vector autoregressive process of an infinite order, ok. So, this is nothing but VAR infinity. So, in short, a VMA1 process can be written down analogously as a VAR infinity process. So, a vector autoregressive infinite process can be rewritten as a vector moving average process of order 1.

And the last thing we will deal with in today's session is we can sort of get the autocovariance matrix of the process and so on. So, I can write down the autocovariance structure. By this expression, so gamma L equals the expectation of YT YT plus L, and then again by taking some cases, so let us say what would happen if L is 0, what happens if L is 1, or what happens if L exceeds 2, and so on and so forth. I can put forward some expressions for writing down the autocovariance structures of the VMA(1) process. So, sigma plus theta 1 sigma theta 1 prime if L is 0, then sigma into theta 1 prime if L is 1, and then 0 if L is greater than or equal to 2.

Now, in the next session—the last session, which will be theoretical this week— I will try to take up one example, and towards the end of the next session, we will also elaborate on or sort of give you lots and lots of examples where you can sort of apply, let us say, either a VARMA process, a VAR process, or a VMA process in a practical sort of situation.

Thank you.