Time Series Modelling and Forecasting with Applications in R

Prof. Sudeep Bapat

Shailesh J. Mehta School of Management

Indian Institute of Technology Bombay

Department of Applied Statistics/Finance

Week 11

Lecture 54: Markov Switching Models

Hello all, welcome to this course on time series modelling and forecasting using R. So, throughout this week just to give you again a very quick refresher. So, we have been talking or rather focusing more on non-linear time series processes and today's session would be the last session from a theoretical point of view. So, of course, we have one more session which would be a practical session in R. But so far I will just sweep you all of you through what we have covered so far this week.

So we started building upon the motivation behind why should one switch from a linear time series structure to some sort of a non-linear time series structure and then we gave an idea or rather an introduction about regimes. So in this entire literature there is a very profound kind of suitability of this term regime and again again if you are not confident about any of the terminologies or definitions again you can always go back to some of the previous lectures or previous videos in this week. So, again what do you mean by regime? So, regime is nothing but different states within a time series process and one strong assumption is that each of these different states

on each of these different regimes should have some dynamically different characteristics ok and throughout this entire week. So, over the last 3 sessions I have been giving some visual explanations as well right. So, let us say if you have an overall non-linear kind of a structure, but and then let us say you have some threshold values let us say 2 threshold values or 3 threshold values. So, depending upon the number of threshold values you have we are kind of dividing the overall time series into those many regimes. And one again a strong assumption is that even if the overall time series becomes non-linear, but within each regime or within each phase or within each state, I can apply some sort of a linear time series process.

Let us say AR, ARMA, ARIMA, or something like that. And in this period, we talked about the very first kind of a non-linear time series model, which is the TAR model, or threshold autoregressive time series process, OK. And then, towards the end of, let us say, the last session or the last-to-last session, we talked about multiple other extensions of the TAR model, let us say, the SETAR model, which is called the self-exciting TAR, or MTAR, or STAR model, right. So, all these are kind of—they sort of point towards different characteristics within the non-linear time series framework. And then, sort of pointing to different ideas about, let us say, if you have, let us say, a different transition function, then what would happen.

So, again, if you remember, we discussed a number of different kinds of regimes, right? And again, if you remember vaguely. So, we talked about three different kinds of regimes. So, the first one was based on a threshold value; the second one was based on a smooth transition function. So, if While a time series switches between, or rather over, a threshold value, if its switch is sort of smooth in a gradual manner, then we end up being in the second case.

And the third is the focus of today's session right here. So, if you see the title here, we will talk briefly about something which is called the Markov switching models. And again, this was seen earlier as well as a third category of kind of a regime where you have a probabilistic kind of a view. So, rather than the transition function being smooth or you having some sort of a threshold value, but what would happen if you bring in some idea of probability. So, let us say you have a probability transition matrix, sort of, right.

Then that category of models or that category of regime techniques is called Markov switching models. So, hopefully up to now there should not be any confusion. So, again I am pretty sure that almost all the topics here might be new to many of you, and again, as mentioned or rather given as a small disclaimer towards the beginning of this week's lectures, this entire area about non-linear time series processes is not discussed often in a routine time series course. So, hence you would want to put a bit more attention to each and every, let us say, model structure of TAR or model structure of STAR. How is STAR different from TAR?

Right. What is the advantage of, let us say, STAR over TAR? So all these small things, if you keep in mind, then hopefully when you get into the practical session, the next session, it should not be any problem. All right. So we start with the title today, which is Markov Switching Models.

Again, as always, a short introduction about, let us say, Markov switching models. So, again in short, they are called MSMs or Markov switching models, and these are a class of non-linear time series models. So, again, the idea is that we are trying to capture non-linearity within the time series that captures regime-switching behavior by assuming that the transition between regimes is governed by some sort of a hidden Markov process. So, I will say that this is the most important differentiating point when it comes to, let us say, Markov switching models or STAR models or rather TAR models, etc. So, again, just to repeat, the differentiating point between MSMs is that the transitioning between the regimes by any practical time series is governed by something called a hidden Markov process.

And the second point is, unlike the TAR models or threshold autoregressive—or rather, the SETAR model or the self-exciting TAR models—where regime changes depend on observable variables, right. So, we have seen this in the past as well: if you take, let us say, a TAR model or a SETAR model, then all the regime switches or all the regime transitions depend on some observable variables. And these observable variables are nothing but the threshold values, ok. And again, we kind of assume that usually we can take the threshold values to be some sort of lagged values of the time series themselves. So, let us say if you are taking y_t as a time series, then I can have something like y_t minus 1 to be the threshold value, and so on and so forth.

So, I will write down etcetera here. So, again, it will not be y_t minus 1 every time; it can be some lagged value of the actual time series y_t itself, ok. But the underlying fact between these two models—TAR and SETAR—is that the regime changes depend on observable variables. And on the other hand, MSMs allow the regime to switch probabilistically based on some unobservable state variables. So, I think, just to again—just one last time—press on the differentiating aspect: if you take up a TAR model or any of its extensions, let us say STAR or ACTAR etcetera, then all the regime switches depend on observable variables.

So, I will highlight this: whereas, if you come to Markov switching models, then all the regime switches occur probabilistically based on some unobservable state variable. So, in MSMs, the regime switches happen using an unobservable state variable, whereas in the usual TAR models or SETAR models, the regime changes or the regime transitions depend on observable variables. Make sense so far? Okay, so now the next thing is: what are the key features of the Markov switching models, right? So, key features of the MS model.

Again, as always, so probably all these pointers you must have seen even in the last session when we talked about let us say key features about the other models. So, TAR model or STAR model, SE TAR model, etc. So, the first one is regime switching. So, the time series can exist in multiple regimes. For example, let us say high volatility, low volatility, right, etc., each with its own set of parameters.

So, there has to be some regime switching which has to be performed underneath the practical time series, okay. Now, here one extra feature and again the most important one is the Markov process. So, what do you mean by that? So, all the regime transitions are governed by a first order Markov process meaning the current regime depends only on the previous regime. Now, this is one very useful characteristic of any Markov process.

So, again I am not sure that how many of you are familiar with let us say Markov process or in general the Markov property. So, the Markov property means that let us say if you are sitting today let us say y t. So, information of y t only depends on one past value let us say y t minus 1 and that is all. So, this is called as a Markov property and if you create an entire time series structure or rather a stochastic structure using this property that process would be called as a Markov process. Similarly, here we have an underlying idea about a Markov process inbuilt in the MS models. And again, how exactly?

So, all the regime transitions are governed by a first-order Markov process. What do you mean by that? So, meaning the current regime depends only on the past regime or one previous regime. Now, the third important aspect here is the hidden states. So, you have to assume that some states are hidden, that the regime is a latent or rather unobservable variable inferred from the data.

So, here I think the most important or rather the key differentiating point is that all these regimes are sort of brought from some unobservable variables which are called latent variables, and hence all these regimes or all these states are called hidden states. And lastly, flexibility as always, so suitable for capturing abrupt or smooth transitions dependent on the state transition probabilities. So, depending on what the state transition probabilities are, we can actually govern if the transition would be smooth or abrupt, and so on and so forth, ok. So, hopefully, if you understand the key features of the MS models, I think pointers 2 and 3 are sort of differentiating from, let us say, either TAR models or STAR models, etcetera. And again, just to repeat one last time, is that all the regime transitions are governed by some unobservable variables.

Or rather latent variables or some probabilistic kind of manner, ok. All right. So, now, we will pay attention to the actual model formulation. So, a simple Markov switching autoregressive model. So, in short, MSAR.

So, again, the key here is that such a model also consists, or rather is based on, an autoregressive kind of structure. As before, right? All the models we discussed so far sort of depend on some autoregressive structure, which is inbuilt in that, right? Similarly, here, a simple Markov switching autoregressive, in short MSAR, model with k regimes. So, let us see: the number of regimes is k, and it is defined as follows.

$$y_t = \mu_{s_t} + \phi_{s_t,1} y_{t-1} + \phi_{s_t,2} y_{t-2} + \cdots + \phi_{s_t,p} y_{t-p} + e_t$$

So, y_t equals mu_s_t plus phi_s_t comma 1 y_t minus 1 plus phi_s_t comma 2 y_t minus 2, and so on and so forth, up to phi_s_t comma p y_t minus p, plus some random error. Now, again, pay attention to all the notations. So, firstly, what do you mean by S_t? So, S_t is the key notation here, which is nothing but the latent state or latent regime at time t. So, again, S_t is the differentiating notation if you compare this model structure with, let us say, TAR model structure or STAR model structure, etcetera. So, the extra thing here is S_t, which is nothing but the latent state or latent regime at time t. And, by the way, again, just to remind all of you that this S_t, being a latent state, is unobservable.

So, all these S_ts are unobservable here, and this S_t can take values between 1, 2, 3, 4, up to k, because how many regimes do you have? You have k regimes, right? And, similarly, mu subscript S_t is a regime-specific mean. So, corresponding to that regime, how much is the mean is characterized by or given by mu_S_t. Make sense?

And lastly, phi st comma i are as usual the AR coefficients within the regime st, ok. And ET is the error term with variance sigma square st. So, let us say we are assuming that this ET is a random error term having this as the variance, ok. So, hopefully model structure should be understood otherwise if you do not understand some of the notation as to what is going on in this MSAR kind of a model then probably it might not be very helpful in the later stage. So, again if you want I can repeat one last time is that ST is the latent state or latent regime at time t taking values between 1, 2, 3, 4 up to k.

Mu ST is a regime specific mean, phi ST, I are all the AR coefficients in the regime ST and ET is the error term with assuming that the variance is sigma square ST. So, this is broadly the model structure of the MSAR modeled with K regimes. Now, the second important ingredient is we have to define some sort of a transition probabilities, right. So, this entire idea about regime switching underlying this MSAR model or in general any Markov switching model is controlled by a state transition probability matrix. which is given by capital P here, right.

$$P = \begin{bmatrix} p_{11} & \cdots & p_{1k} \\ \vdots & \ddots & \vdots \\ p_{k1} & \cdots & p_{kk} \end{bmatrix}$$

So, this matrix that you see which is capital P is called as a state transition probability matrix and again I am very sure that we have discussed this in some of the past lectures this week, right. So, again if you vaguely remember that we had put down a similar matrix and then. So, if you remember while discussing the three kinds of regimes, right and then the third kind was exactly this one which is Markov switching regime, we had kind of introduced the notation about a state transition probability matrix, So, similarly here the only idea is that we are kind of extending this to more than 2 up to k. Because we are assuming that number of regimes are k right. So, again just to press on that fact that number of regimes equals k. So, this matrix would be a k cross k matrix ok.

So, if you want the order of the matrix, it would be k cross k. Now, what exactly are all the individual items in this matrix? For example, P11 or, in general, Pij. So, what do you mean by Pij? Where Pij is nothing but the probability that St is j while St minus 1 was i. So, the current regime value is j, but the last regime value or one past regime value was i.

So, in short, it means the probability of transitioning from regime I to J because the last regime was I, the current regime is J, and how much is the probability that the time series would transition from, let us say, regime I to regime J, basically, OK. So, I am pretty sure that understanding this state transition probability matrix should not be difficult at all, right? Because all these are sort of probabilities, and each probability has two subscripts, right, something like i j. So, think that i is the last regime state, j is the current regime state, and how much is the probability that it will make the transition from regime i to regime j. That is all. And, by the way, one very important property is that each row of P sums to 1. So, if you take a sum of any of the rows, let us say something like this: summation P and then i j, or rather 1 j, let me write down 1 j, and then j going from 1 to k, this has to be 1 for all j, for all j. Make sense?

And then, if you take a sum over any of the rows, it has to be exactly equal to 1. Alright, so now we will pay attention to the estimation part, right. So, how does one estimate? So, again, we can make use of some of the standard techniques, let us say, create the likelihood function. So, the joint likelihood of observed data and hidden states is maximized.

So, once you create that likelihood function, then how do you maximize that? How do you get hold of the MLE, etc.? But here it might require marginalizing over the latent states. So, you would want to find out the marginal distribution over the latent states. This is an added step here, which has to be performed.

The second kind of estimation technology—or rather not technology, but method—is the expectation-maximization algorithm, or in short, the EM algorithm. And again, this EM algorithm is widely applicable, not just in time series but across any other statistical literature as well, right? So, the EM algorithm contains two steps. The first is called the E-step, which calculates the expected values of the hidden states given some set of parameters, right? So, this is the expectation part, or the E-step, right?

And the second step is the M-step, or the maximization step. So, the M-step updates the parameters to maximize the expected likelihood. So, rather than writing down a joint or some sort of overall likelihood function, what we do is run this algorithm in two levels. The first level is the E-step, which finds out the expected values of all the hidden states given some observable set of parameters. And once the E-step is done, then the next step is the M-step, which updates the parameters to maximize the expected likelihood.

And then the third kind of estimation technique is filter and smoother. So, this is slightly technical. Again, we will not go into details as to what exactly you mean by filter or what you mean by smoother, but just a couple of pointers here that in literature these two filters exist. So, the Hamilton filter estimates the probability of being in a particular state at time t, while the Kim filter smooths the probabilities to improve the state estimation. So, hopefully the idea about estimation should be clear as to what the different techniques are that one can use to estimate the underlying parameters of a Markov switching kind of model.

So, it could be either using a likelihood function or the EM algorithm or using a filter and smoother kind of technique, etcetera. Now, what exactly are the advantages of MS models, right? Again, as always, flexibility. So, it can handle both abrupt and gradual transitions between regimes. Second is state dependence.

So, what do you mean by that? So, it incorporates dynamics specific to each regime, improving its model accuracy, okay. So, it incorporates dynamics specific to each regime because, again, as discussed a number of times in this entire week, each regime is entirely different or could be entirely different, having different sets of dynamics, okay. So, state dependence means that it incorporates dynamics specific to each state or specific to each regime, improving the overall model accuracy. And thirdly, latent regime modeling, which infers unobservable states that drive the system, okay.

And similarly, on the other hand, what could be the challenges in the MS model? So, again, computationally complex, right? So, this MS idea is not very straightforward because you have to form an idea about the transition matrix, and it involves some sort of probabilistic view, okay? So, computational complexity exists, which means that estimation involves maximizing a high-dimensional likelihood function. So, it may turn out that the overall likelihood function, which has to be maximized, is a high-dimensional likelihood function.

So, containing lots of parameters or, you know, lots of levels, etcetera. Secondly, the other challenge would be model selection. So, choosing the number of regimes firstly and lag structure requires care. So, how many regimes should you have, and up to what lag in the path should you go? So, the lag structure and the number of regimes require some care.

Thirdly, interpretability. So, regime probabilities must be carefully interpreted. So, when you create that capital P matrix containing all the transition probabilities, then each and every switching probability, or rather regime probabilities, must be carefully handled or carefully interpreted. And lastly, overfitting. So, adding too many regimes can lead to overfitting, as seen in the earlier models as well.

So, these are some challenges when it involves, when it comes to MS models. So, I will say that model selection, computational complexity and interpretability are slightly on the newer side. Well, of course, overfitting is always there. So, if you add in too many regimes, not just in the MS kind of a model, but even in TAR models or STAR models, then that can lead to some overfitting. Now, what we will do is I will show you one particular numerical example.

So, since we are almost through with covering all the theory, so I will talk about one numerical example and then this would be helpful in understanding as to how do you create that transition matrix and how do you fix some of the values, etc. So, small numerical example demonstrating a Markov switching autoregressive or other MSAR model containing two regimes, ok. So, let us say number of observations is fixed to 300, the mean vector is fixed to 1 comma minus 1. So, since you have two regimes, then each vector should contain two values, ok. Then the phi vector is 0.8, 0.5, this contains the AR coefficients and the standard deviations of the error terms is 0.5 comma 1.

And this is the probability transition matrix. And again one characteristic is that if you sum over any of the rows it should be 1. So, 0.9 plus 0.1 is 1 while 0.2 plus 0.8 is also 1. Now, by the way all these values are kind of hypothetical values. So, there is no backing to it.

So, this is just some random values. But then what? So, let us say once you fix all these ideas, then the next step is estimation, right? So, estimation interpretation or rather estimated parameters. So, the mean value, which is mu, the AR coefficients given by phi, and variance given by sigma square for each regime, has to be estimated.

Also, the transition probabilities, let us say Pij, that show the likelihood of switching between regimes. So, all these are considered to be parameters which have to be estimated. So, let us say mu, phi, sigma square, and also Pij for each state. And then, secondly, hidden states. So, smooth probabilities indicate the likelihood of the time series being in each regime at a given time.

So, you have to carefully address the idea of hidden states or estimation in general, okay? Alright. So, let us say if you run this entire code in some software, let us say R or Python, wherever, and then you try to estimate all the parameters given the sets of input values which we saw in the previous slides. Then this is exactly one set of estimation that the software throws at you. So, inside regime 1, what happens is mu 1 is estimated to be 0.98, phi 1 is estimated to be 0.81, while sigma 1 is estimated to be 0.52.

And similarly, in regime 2, the mean value is -1.02, phi 1 is 0.48, and sigma 1 is 1.05. Now, again, just to compare these estimated values with the actual values, if you remember vaguely what values we fixed earlier. So, our mu 1 was something like 1, if I am not wrong, and mu 2 was something like -1. By the way, here we have a small typo. So, this should be mu 2. They should be phi 2, and they should be sigma 2, basically.

So, my mu 2 was -1, and my phi 1 was—again, let me go back a couple of slides. So, my phi 1 was 0.8, phi 2 was 0.5, and again, you can clearly see that the estimations are really close to the fixed values. So, phi 2 was 0.5, while lastly, my sigma 1, if I am not wrong, is 0.5, and my sigma 2 was basically 1. And similarly, transition probability. So, these are the estimated transition probabilities.

So, P11 is 0.89, P12 is 0.11, P21 is 0.19, while P22 is 0.81, and each one of these is sort of close to the fixed values in that probability matrix, which is capital P. So, the summary of the output is giving us a very good indication that the estimation has been performed quite well here. Now, the output interpretation. So, the parameter estimates—mean, the AR coefficients, and variance sigma squared for each regime—and transition probabilities between the regimes. So, the outputs give you these things, right?

So, parameter estimates of all these ideas. Secondly, plot of smooth probabilities. So, if somebody is interested in, let us say, plotting the smooth probabilities, it would show the likelihood of being in each regime over time. Okay. So, one can actually plot the different transition probabilities.

It would show the likelihood of being in each regime over time. So, what is the probability of, let us say, switching from one regime to another or from the second regime to the third, etc.? And thirdly, diagnostics. So, validate the model by checking the residuals. So, once you do the estimation part and once you fix some model on the parameters, on the data, and so on and so forth, then the last step has to always be diagnostic checking.

So, validate the model by checking the residual performance. So, now we will very quickly pay attention to some other possible interesting extensions, right? So, the first kind of extension is called bilinear models, and then how are the differences? So, let us say, incorporate products of past observations and some noise to handle non-linearities, and such bilinear models have the capability of including some interactions.

$$y_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \sum_{k=1}^{p} \beta_{jk} y_{t-k} e_{t-j} + e_t$$

So, here if you see, we are sort of multiplying these two coefficients, right? So, beta j k multiplied by y t minus k. So, you are sort of bringing some interactions in the time series structure. And what are the applications? Let us say, used for time series with interaction effects, such as finance or environmental data, etc. So, one extension is the bilinear model. Another extension could be the non-linear ARX model.

So, what do you mean by that? So, the model's output is a non-linear function of past inputs and outputs. So, this could be treated as you have a time series which runs like that in terms of yt. And on top of that, if you have some other exogenous inputs as well. Or some other independent variables which are given by Xt's, and then you have a function combining Xt's and Yt's.

Such a model is called a non-linear ARX model. So, by the way, ARX stands for autoregressive exogenous, right? So, X stands for exogenous inputs. So, on top of the autoregressive structure, if you have some exogenous inputs as well, then we can have a model which is called a non-linear ARX model, okay. So, in this case, F can be a polynomial, neural network, or some other non-linear function.

And application-wise, it is seen in control systems, dynamic system modeling, etc., okay. The third one could be a neural network-based model. So, let us say a nonlinear autoregressive neural network, or in short, NAR/NN, uses a neural network to model the autoregressive relationship. By the way, the entire next week will be spent on, let us say, machine learning and time series integration. So, there again, we will talk about, let us say, the characteristics of, let us say, the NAR model or NAR/NN model, etcetera, okay.

Flexible and powerful, especially for complex nonlinear dependencies. So, one can go the neural network route if you want to perform some complex or capture some complex nonlinear dependencies. On the other hand, LSTM. So, again, I am pretty sure that most of you might have heard about this term. So, long short-term memory handles sequential dependencies and long-term memory in nonlinear time series.

Other possible interesting extensions could be, let us say, smooth transition GARCH or rather ST-GARCH. So, it extends GARCH by allowing regime switching in the variance equation. So, we have studied GARCH extensively in the last week. So, can we extend a GARCH model by bringing in some smooth transition function? So, it is called the ST-GARCH model.

Or, on the other hand, let us say hidden Markov models (HMM), which capture switching dynamics in time series using discrete latent states. Then, functional time series, let us say, are useful for high-dimensional or infinite-dimensional time series data, modeling temporal changes in curves or functions. And lastly, wavelet transform models decompose time series into different frequency components and are useful for identifying and modeling non-linear periodic patterns. So, again, the literature behind—or rather within—this non-linear time series modeling and forecasting is vast, right? So, we have discussed a lot of extensions, which are interesting and applicable to many more complex kinds of situations.

So, one should always choose the correct model depending on the intricacies in the underlying dataset. Thank you.