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 53: Nonlinear Model Extensions

Hello all, welcome to this course on time series modeling and forecasting using R. So, again, just to give you a very quick refresher as to where we are standing this week. So, the broad idea throughout this week has been about non-linear time series processes. In the very first session, we gave an introduction, or rather a motivation, as to when or exactly why one should transition from capturing linearity in practical data when it comes to time series processes to non-linearity. And there, we gave some examples as to when one encounters nonlinearity in practical datasets, etc., okay. And just in the last session, we discussed some of the different types of regimes.

So, one very important idea behind capturing nonlinearity is the transitioning of a time series between regimes or between states, okay. And there, we talked about different kinds of regimes. So, smooth transition functions, or from a probabilistic point of view, where we discussed Markov switching models, right? Or the very widely applicable one is called a threshold-based regime, okay. And inside that, we talked about one very particular kind of time series model called the TAR model, or T-A-R, or the long form of that is the threshold autoregressive kind of model. And again, the threshold autoregressive model assumes that, even if you divide the time series into multiple regimes, one can actually encounter or try to model each of the regimes within itself using some linear structure.

So, you have an overall nonlinear structure, but within each regime, if you delve deeper into each of the states or each of the regimes, you can put forward some linear model, let's say an AR kind of structure or an ARMA kind of structure inside each of the regimes. So, this was the motivation for developing or discussing the TAR model. Now, the idea behind today's session this week will be to discuss further extensions of, let's

say, the TAR model or regime-based models, etc. Because the very first model we discussed in the last session, the TAR model, is not a very advanced kind of model. So, as probably many of you might have guessed or might have understood, TAR is kind of dependent on some of the past values of the time series itself because it has an autoregressive component.

But, we did not bring in any of the other components or did not tie down the TAR model with any of the other extensions so far and this would be the basic crux of today's session. So, the title that you see stands for further extensions. So, before we go into detail in any of the extensions probably we will discuss a list of all the extensions of the basic TAR model. So, the title of the slide is extensions of the TAR model. We will start with the very first one which is called as SE TAR.

So, SE TAR is called as self-exciting TAR. What do you mean by that? So, a TAR model where the threshold variable is the time series itself. So, it is called a self-exciting TAR. So, the threshold variable is nothing but some of the lag value of the time series itself.

So, y t minus d. So, let us say y t is the actual time series it may happen or you may sort of assume that let us say y t minus 1. So, just the past value could be the threshold value ok and under these cases we will say that the underlying TAR model would be called as self exciting TAR model. Now, the second kind of a TAR model or second extension could be momentum TAR or in short mTAR and what is the difference here? So, mTAR model uses the difference or momentum let us say delta of y t minus d as the threshold variable.

So, again let us say, so this delta variable is sort of similar to the nabla variable we discussed earlier, much earlier. So, again if you are not very confident about nabla variable, nabla is nothing but the first difference. So, again if you reread the definition of this mtar kind of a model, it uses the difference or momentum, let us say nabla applied on yt minus d as the threshold variable. Then the next one could involve multiple thresholds. So, let us say allows more than two regimes extending the model to multiple regimes with distinct dynamics.

So, again so I will say that multiple thresholds is just a very very easy extension of extending from one regime to more than two regimes or rather two regimes to more than two regimes. So, again just to give you a small example very quickly is that if you have a time series which behaves like that. And here clearly you can spot three regimes right. So, let us say I can draw a line here, I can draw a threshold value here. So, I am sort of

dividing the overall time series into three different regimes where in the first regime you see some sort of a not that volatile structure, but in the second regime you can see highly volatile nature and rather in the third structure you can again see sort of a stationary kind of a behavior or again non volatile kind of a structure ok.

So, multiple threshold means that if you are sort of dividing the time series into more than two regimes, then you have those many thresholds. And the next one is star model. So, again I will say that tar model which is TAR and the star model are kind of the most widely applicable ones. So, full form of star is smooth transition autoregressive which is nothing but a variant where transitions between regimes are smooth modeled using some logistic or exponential transition functions. So, again the star models fall into the second category we discussed in the last session where if you remember we discussed three different classes of regimes.

So, the first one was threshold-based. And then obviously, TAR is one example of that. Then the next one where we brought in the idea of some transition functions, right? And there, the TAR models would come into play, okay? So, smooth transition autoregressive, which is a variant where transitions between regimes are smoothly modeled using some logistic or exponential transition functions.

So, again, I will say that these are some extensions of a basic TAR model. So, if you are not very confident about a basic TAR model, the first thing is you should go back to the last session, get hold of or try to understand the model structure because, again, if you remember, we discussed the model structure in the last session as well, which takes the form of an autoregressive structure, by the way, and all these could be simultaneous extensions of that model structure. Alright, so now moving ahead, I think this is just the exact same thing here again, just a sort of a reminder of what all models we discussed in the last slide. So, let us say SETAR, the second one is MTAR, the third is multiple thresholds, and the last one is STAR or smooth transition autoregressive. So, now the next model we will discuss or put a bit more detail into is the last one in the previous slide: this is the TAR model, this is the smooth transition autoregressive model.

So, as the heading says, the smooth transition autoregressive model, we will try to understand what exactly this is, how this is largely different from the TAR model, etcetera. So, STAR models are a type of nonlinear time series models where the transition between regimes occurs gradually rather than abruptly, as in threshold autoregressive models. So, the transitioning between the different regimes is sort of smooth; it is not

abrupt. This is the whole difference between STAR models and TAR models. So, the transitioning of the time series between regimes across the threshold value happens in a smoother fashion or gradually rather than abruptly, as seen in TAR models.

And what exactly is the motivation behind sort of extending the TAR to star? So star models are useful when the underlying system exhibits smooth regime shifts, making them suitable for capturing dynamics like policy effects, economic cycles or gradual climate changes, etc. So these are some of the examples where star models are or rather prove to be advantageous in a way. Now, what exactly are some of the key features of star model? So, the first one could be smooth regime transition.

So, as discussed a short while back, so the transition function is smoother as compared to tar models. So, one encounters or one sort of sees a smoother sort of a regime transition. So, unlike TAR models where regime switch instantly or regime switch instantly, star models use a slightly smoother transition function to model the gradual changes in the transition or the underlying time series features. So, as the name suggests S star, so the S in the star stands for smooth. So, smooth threshold autolegressive model is an extension of a TAR model where all the transitions happen in a smoother fashion.

Now, the second point is there has to be some transition function, right? So, there has to be some continuous function by the way. So, the transition functions are continuous functions. Some of the examples are logistic or exponential which determines the smoothness and location of the transition. And thirdly, flexibility.

So, combines the simplicity of linear models within the regimes with the ability to model non-linear dynamics across the regimes, right? So, I think we discussed the idea of flexibility even the last session this was one of the key features of the TAR models also and the exact same thing can be seen as a key feature of the TAR models. So, again what do you mean by flexibility? So, flexibility combines the simplicity of linear models within each of the regimes with the ability to model the non-linear dynamics across the regimes. So, again just to list down these are some of the key features when it comes to

the idea of star models ok. All right. So, now, the next next big thing or rather the elephant in the room is defining the model structure of the star model and immediately as you can notice or as you can understand the model structure is not at all easy. So, again one small disclaimer that you are not very confident about any of the notations you might have to visit some of the earlier lectures or rather not go into too much detail and then get bogged down by all the terminologies. So, again a suggestion would be that try to

understand what notations go inside the model rather than the explicit nature of the entire model ok.

But in general a general star model can be expressed as follows. So, let us say y t which is time series itself equals phi 0 plus phi 1 y t minus 1 plus up to phi p y t minus p. Plus, there has to be some transition function, which is given by capital G, as seen in the last lecture as well. So, this capital G is nothing but the transition function, which depends on three things. So, ZT minus D, gamma and C, where ZT minus D threshold variable with lag D, this is the transition function, right?

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + G(z_{t-d}; \gamma, c)(\psi_0 + \psi_1 y_{t-1} + \cdots + \psi_p y_{t-p}) + e_t$$

And then gamma is nothing but the coefficient, and C is nothing but the threshold value. And this entire transition function is again similarly multiplied by a similar-looking structure in terms of different coefficients in terms of psi's. So, psi 0 plus psi 1 yt minus 1 plus all the way up to psi p yt minus p plus et. So, in short, what is happening here is that you have an AR structure here plus a certain transition function multiplied by a slightly different AR structure plus a random error. Where Z t minus d is nothing but the threshold variable with lag d. This is the transition function controlling the regime shift.

Phi i's are the parameters for the base regime, while psi i's are the parameters for the second regime. Make sense so far? So, again, We are trying to explain the model structure in a bit more detail, right. So, hopefully, the model structure would be understood now—that there should be two different structures, one in the first regime and the second in the second regime.

And this is nothing but the autoregressive structure in the first regime, and this is nothing but the autoregressive structure in the second regime, connected through or rather connected via this transition function. Make sense? So, in literature, what are the different kinds of transition functions that one can encounter, right? So, a very famous or rather widely applicable transition function is a logistic transition function, which all of you—or rather all of us—saw in the previous lecture also when we defined the idea of smooth transition functions the very first time, okay. Now, again, just to describe this in a bit more detail exactly as to how the logistic transition function looks like.

$$G(z_{t-d}; \gamma, c) = \frac{1}{1 + e^{-\gamma(z_{t-d} - c)}}$$

So, this is an advantage of smoothly transitioning between regimes as Z T minus D crosses a threshold value, which is C. So, again, I will probably draw your attention to this piece in the function, which is Z T minus D minus C. So, the moment Z T minus D crosses C, we are seeing a transition, basically. And this entire structure takes the form of a logistic kind of function, and hence the name is logistic transition function. So, 1 divided by 1 plus e to the power minus gamma into z t minus d minus c. Now, how exact is the behavior?

$$G \to 0; \quad z_{t-d} \text{ is far below } c.$$

$$G \to 1; \quad z_{t-d} \text{ is far above } c.$$

So, on one hand, if this transition function converges to 0, then we can say that z t minus d is far below c, and when this transition function approaches 1, we can say that z t minus d is far above c. So, rather, the c is nothing but the threshold value here.

So, as described earlier in the same kind of example. So, and then we are sort of comparing the value of z t minus d with c. So, when z t minus d crosses c, we are having a transition, okay. So, whenever the value of this transition function approaches 0. So, on one end, we will say that z t minus d is far below c. So, the transition is not happening. And on the other hand, if G approaches 1 or converges to 1, we will say that the transition is happening because Z P minus D would be far above C.

And these are some of the key applications. So, modeling gradual shifts like economic expansions, etc. All right. So, now, on the other hand, the second kind of a very widely applicable transition function is nothing but the exponential transition function, a slightly different function as compared to the earlier one. So, in the last slide, we talked about logistic, and in this slide, we will spend some time on getting into details about the exponential transition function. And firstly, one should understand that where exactly is this applicable or advantageous. So, useful for symmetric transitions around C. Now, again why symmetric? Because you have a square term here. So, please try to understand that in the earlier logistic case, we did not have a whole square.

$$G(z_{t-d}; \gamma, c) = 1 - e^{-\gamma(z_{t-d}-c)^2}$$

So, it could be asymmetric on either side. But the moment you take a square, so for example, minus 2 square is also 4 and 2 square is also 4. Right. So in and around C, it would be a symmetric kind of a structure. Right.

And hence, the description says that exponential transition function is useful for symmetric transitions around C. So again, from a visual point of view. So what do you what do you mean by this is that let us say if you have this is the threshold value, which is C here. And let us say the time series behaves like that. And then again, it transitions as it crosses C and probably behaves like that. But if you see closely around C, you can have a symmetric nature here.

So, symmetric means a sort of mirror image around C. Such transition functions are useful for symmetric transitions around the value C. And this is the transition function model structure. So, 1 minus e to the power minus gamma and then z t minus d minus C whole square. A couple of characteristics are: whenever g is 0, it means that z t minus d is exactly sitting at C. Whenever g approaches 1, then z t minus d moves away on either side of C. So, it could be either in the second regime or in the first regime. So, these are some of the intrinsic characteristics when it comes to the exponential transition function. And now, lastly, application.

So, systems where deviations from C are symmetric. For example, let us say oscillatory dynamics, etc. So, if you see a pendulum, it oscillates uniformly in a symmetric kind of nature. So, if you sort of put different threshold values in between the oscillations, then around any of the threshold values, you can see symmetric behavior of the pendulum or the oscillatory dynamics. So, again, before we proceed, hopefully, the idea behind developing or discussing these different transition functions must be clear, right?

So, what is the advantage of logistic, what is the advantage of exponential, etc.? So, each and every transition function has its own pros and cons, basically, right? So, which is the one capable of handling, let us say, symmetric transitions, or which is the one capable of handling non-symmetric or other asymmetric transitions? So, all these things should be kept in mind, rather, okay? Okay, so now the next thing we will discuss is steps in model building.

So, the very first step is specify the threshold variable. So, this is very very very important. So, in this case the threshold variable is Z T minus D. So, this is often a lagged value of the series which is Y T minus D itself or it could be some external variable. So, definition wise and notation wise, this is the threshold variable denoted by Z. So, Z T

minus D, which is nothing but a lag value of the series. So, this is the series Y T minus D itself or it could be some external variable also.

Secondly, estimate linear model. So, fit a linear AR model as a benchmark to test for non-linearity. This is the second step. So, once you specify the threshold variable, second step is estimating some linear model. So, one should always fit some sort of a linear let us say autoregressive structure as a benchmark to test for non-linearity.

So, again why is the second step important? Let us say you try to model that using a simple linear sort of an autoregressive model and then the model is not performing very good. Then this would be sort of a benchmark model to test for further non-linearities which are hidden in the model. And obviously, third step is test for non-linearity. So, conduct linearity tests, example Lagrange multiplier or LM test to justify using star model, ok.

So, these are some of the initial steps when it comes to model building. And the fourth step would be selecting the transition function. So, since the star model revolves around applying or sort of picking the correct transition function, selecting the appropriate transition function is at the more important. So, choose between either logistic or exponential transition functions based on the application in hand.

So, should you go with a logistic transition function or should you go with exponential transition function, this has to be decided beforehand, ok. And the next step is estimating the parameters. So, use of non-linear optimization techniques, let us say MLE technique or non-linear leaf squares to estimate all the parameters. So, again what are the parameters here? Let us say gamma is the important parameter, C is the important parameter and all the AR coefficients are the other parameters.

So, gamma C and whatever number of AR coefficients that the model would have. And then lastly, model diagnostics. So, once you estimate the parameters, try to fit some star model, then check the residuals for all the problems or the idea is that there should not be any problem. So, autocorrelation, changing variance and adequacy of the smooth transition. So, all these ideas come under the diagnostic checking heading.

Now, the second sort of extension is called self-exciting TAR models or self-exciting threshold autoregressive models, or in short, SETAR models. So, these SETAR models are a specific type of threshold autoregressive or TAR models in which the threshold variable is a lagged value of the time series itself. Again, I think this first point is exactly

similar to what we had in the STAR model as well, but again, we will try to understand how SETAR is slightly different from STAR, okay? The second point is this self-referencing nature gives SETAR models their name and makes them effective for capturing regime-switching dynamics based on the history of the series itself. So, one very big advantage is that since the threshold value has to be some of the lagged values of the time series itself, you have some sort of self-referencing nature in that, right?

So, let us say yt is the running time series, and for some reason, let us say if you assume yt minus 2 to be the threshold value. So, you have some sort of self-referencing nature. So, this is the threshold value; this is the current progression. So, you are basically comparing the current progression with the threshold value, which is nothing but the past value of the time series itself. And the advantage here is that the self-referencing nature gives SETAR models their name and makes them effective for capturing regime-switching dynamics based on the history of the time series itself.

And these are some of the key features of the SETAR model. So, let us say regime-specific dynamics. So, the time series switches between regimes based on the value of its own lagged observations. Then, threshold determination. So, the lagged value, let us say y t minus d, is compared to one or more threshold values, let us say gamma, to define regimes.

So, how do you sort of fix on one threshold value? So, one compares some of the lag values of the series itself, let us say yt minus 1 or yt minus 2, etc. And all these are compared to some standalone threshold value, let us say gamma, to sort of specify the regimes. Thirdly, as always we have seen even earlier, piecewise linearity means the model is linear within each regime but non-linear as a whole due to regime switching. And lastly, again, we have seen this number of times: flexibility suitable for processes with abrupt regime changes or asymmetric behaviors.

So, these are some of the key features when it comes to the S-E-TAR models. Now, again, the exact model formulation. So, again, one small disclaimer yet again: the model formulation is not very easy, but since all these models sort of have the inbuilt autoregressive structure, understanding each and every model is not that difficult, right? Because within each regime, one can clearly see that one preserves the autoregressive structure which is inbuilt. So, the AC TAR model can be easily expressed as follows. So, again, you can see two levels.

$$y_t = \begin{cases} \phi_{1,0} + \phi_{1,1} y_{t-1} + \cdots + \phi_{1,p_1} y_{t-p_1} + e_t, & if \ \ y_{t-d} \leq \gamma \\ \phi_{2,0} + \phi_{2,1} y_{t-1} + \cdots + \phi_{2,p_2} y_{t-p_2} + e_t, & if \ \ y_{t-d} > \gamma \end{cases}$$

So, the first level is phi 1 comma 0 plus phi 1 comma 1 all the way up to phi 1 comma p 1 y t minus p 1 plus random error, and this is true whenever the value is less than or equal to gamma, and it switches to a different value. Let us say the difference is from 1 to 2 in the second level whenever y t minus d is bigger than gamma; then the second level has to be And again, just to specify the notations: y t is the time series itself at time t, y t minus d is the threshold variable with lag d. And here, clearly, as one can see, y t minus d is nothing but the lagged value of the time series itself. Gamma is the threshold value, and phi i comma j is the AR coefficient for regime i. And this is one extra thing if you compare this to TAR models: P1, P2 are the lags in the AR processes for the respective regimes. So, here, what might happen is that if you have two different regimes, then within each regime, I am trying to fix the AR process having different lags.

So, P1 and P2. And ET, as always, is a random error. So, this is the overall model structure when it comes to SETAR models, and then the whole idea here is that the threshold value is taken to be generally one of the lagged values of the series itself, and that threshold value is compared to some standalone value, which is gamma. Now, steps in model building. So, again, as always, let us say determining the threshold variable.

So, typically, let us say y t minus d, which is nothing but a lag value of the series, is used to determine the threshold variable. Secondly, the more important point when it comes to SETAR models is choosing the number of regimes. So, how many regimes should one have? So, commonly, two regimes are considered, let us say k being 2, but models with more regimes where k is bigger than 2 can also be specified. So, it could be either two regimes or slightly more than two, let us say 3 or 4, but again, as discussed, let us say even in the last session, that one should not go too overboard in deciding the number of regimes.

So, one should not have too many regimes, which might lead to overfitting. The third point is, again, estimating the threshold. So, let us say use methods like grid search to find the threshold value, let us say gamma, which minimizes some sort of an information criterion, let us say AIC or BIC, etc. So, one can again invoke the ideas of AIC or BIC to find out the exact point where the threshold should be placed. Next, as always, estimating the parameters.

So, fit separate AR models within each regime using let us say OLS technique or other estimation techniques and then test for non-linearity. So, perform statistical test such as let us say Hansen's test to confirm the presence of regime switching and then diagnostics. So, validate the model by checking residuals for autocorrelation and performing any of the out of sample forecasting. Now, what exactly the advantages of SETAR model? So, interpretability.

So, the piecewise linear structure is easy to understand and interpret regime specific insights. So, second point is this that it sort of provides separate models for different regimes offering richer insights and again as always flexibility that handles processes with abrupt or threshold dependent changes effectively. But then at the same time, what could be the challenges in SETAR models? So, let us say threshold estimation. So, accurately estimating gamma can be computationally demanding, overfitting.

So, using too many regimes or parameters can lead to overfitting. By the way, please change this to gamma. It is a small typo here. This should be threshold value gamma. And lastly, boundary effect.

So, sparse data near the threshold value can complicate estimation. All right. And all of these one last slide is what exactly are a few extensions of the SE-TAR model itself. So, let us say SE-TAR with multiple thresholds. So, instead of just one threshold I can allow for more than one thresholds or rather more than two regimes right.

Second one could be M-TAR or momentum threshold autoregressive model. So, uses a change in the series let us say nabla yt as a threshold variable. Third one could be CS ATAR or continuous ATAR, a variation where the regime transition is smoothed using continuous functions. And lastly, seasonal S-ATAR model, which accounts for seasonality by incorporating periodic thresholds. So, probably as of now we are done with three sessions of this week and then you might have gotten some important ideas as to the transition between linear time series models to non-linear time series models.

All the different kinds of non-linearity when it comes to non-linear time series models, let us say TAR models or STAR models or SETAR models and all these are extensions of SETAR, right. So, the key idea is that if you want to sort of capture the changing dynamics under any of the practical time series, then obviously modeling that using a linear time series may not be the good thing. So, one would require or one would sort of have to make the shift to capturing the non-linearities in the underlying practical data. Thank you.