Time Series Modelling and Forecasting with Applications in R

Prof. Sudeep Bapat

Shailesh J. Mehta School of Management

Indian Institute of Technology Bombay

Week 05

Lecture 24: Double and Triple Exponential Smoothing

Hello all, welcome to this course on time series modeling and forecasting using R. Now again, essentially just to quickly revise where we are at, so this week we are focusing more on the idea of forecasting. So initially, we started this week of forecasting by explaining the idea of forecasting, as to what is the difference between, let us say, forecasting, prediction, or estimation, etc. And in the very first lecture of this week, we talked about a particular forecasting technique called minimum MSE forecasting applied on a basic kind of time series model which was ARMA. And in the last lecture, we talked more about some smoothing techniques.

So, when can one apply some smoothing techniques, let us say simple moving average or exponential moving average, right? And now in this lecture, we will kind of extend the idea of smoothing techniques, okay. So, by the way, all these extensions or all these alternative structures are kind of improving upon, let us say, a basic kind of smoothing technique, for example, SMA or EMA, okay. Now first thing, just to very quickly revise what you mean by smoothing, is again let me redraw the graph we drew last class. So let us say if you have something like this kind of structure where you have lots of random fluctuations, then probably no fluctuations in the middle, and again probably towards the end you have some noise, right.

So one can clearly see that forecasting such a graph or forecasting such a structure may involve some challenges, right, because it contains a lot of noise or a lot of unwanted kind of repetitions or unwanted kind of replication of the data, especially in this part and that part, right. So, applying proper smoothing techniques is helpful in kind of smoothing the noise or smoothing the unwanted random fluctuations. So, once you apply, let us say, either SMA or EMA or probably double exponential smoothing or triple exponential smoothing, then the idea or the goal is to obtain a graph which would look much more uniform or much more smoothed out. So, probably something like this. Now forecasting such a graph as compared to such a graph is easier, of course.

So now, the first extension we will talk about in today's lecture is called double exponential smoothing, or it is also called the Holtz method. And then, all these are kind of properties or descriptions of that Holtz technique. So, when actually do you use such a technique as compared to EMA or SMA? So, we will study that. So, double exponential smoothing is used for forecasting the time series when the underlying data has a linear trend but no seasonality.

So, I think this part is the most important one: when the underlying data contains some linear trend or some trend, but then no seasonal patterns or there is not any kind of seasonality in the data, then one can actually apply the Holtz technique or double exponential smoothing. Now, by the way, this technique is also called Holtz trend corrected or second-order exponential smoothing, right? So, the first order is the standard EMA that we studied in the last lecture, and then if you want to involve a trend component also on top of no trend or simply smoothing the data using some EMA, then this is an extension of that technique, which is also called Holtz trend corrected or second-order exponential smoothing, okay? So, essentially, here what is the extra thing that we are doing is we are also introducing a term to take care of the trend present in the time series. Now, again, remember the data should contain some trend but no seasonality or no seasonal pattern.

So, such a double exponential smoothing technique, when you kind of go for forecasting using such a technique, is kind of capable of capturing either an increase or decrease in the linear trend in the underlying data. Alright, now what exactly are the model aspects of it, or how do you write down the actual model? So, we will see. So, firstly, as always, if the value of t is 0, then we have this kind of a trivial situation. So, S0 happens to be Y0 to start the process, right, but for any other value which is bigger than 0.

Now, here we have two equations. So, can you see that? So, if you compare this set of equations to the EMA equation. So, in the EMA equation, we only had the first one where you had alpha. But here, since we are also involving some trend idea, we want to capture the trend part with a different kind of parameter, which is beta.

So, alpha is the same that we studied in the last lecture, which is present in the EMA also, but then we have this extra beta parameter, which kind of captures the trend aspect. So, the first equation is exactly similar to what we saw in the EMA kind of instance, that S t equals some weighting parameter, which is alpha into y t plus 1 minus alpha into S t minus 1 plus b t minus 1. So, since you are involving two smoothing series here, one in terms of S t and the other one in terms of b t. So, we have to involve both here. So, S t minus 1 plus b t minus 1, but nevertheless, the first equation looks similar to what we had in the EMA, right?

$$\text{For } t = 0,\ s_0 = y_0.$$

$$\text{For } t > 0,$$

$$s_t = \alpha y_t + (1-\alpha)(s_{t-1} + b_{t-1})$$

$$b_t = \beta(s_t - s_{t-1}) + (1-\beta)b_{t-1}$$

Where,

b_t is the best estimate of trend at time *t*.

$0 < \beta < 1$ is the trend smoothing factor.

So, you had some weighting parameter alpha applied on the current observation y t plus the remainder, which is 1 minus alpha applied on the recursive structure of the model. Now, here, what exactly is B t? So, we have another idea here, which is kind of extra, like we discussed earlier. So, B t kind of captures the trend part. So, what exactly is B t?

So, Bt is some beta applied to the difference of St minus St minus 1 plus the remainder, which is 1 minus beta applied to its recursion. So, Bt minus 1. Now, of course, I understand that the equations are kind of involved here. But then try to understand the fact that you have two things to control here. So, one is the weighting parameter, which is alpha, similar to EMA, and the extra thing we require to control is the trend aspect.

And hence we require two equations, and then both of them are recursive in their own sense. Where Bt is the best estimate of the trend at a particular time t. So, this Bt is the best estimator of the trend at time t, and this extra parameter beta, which would obviously again lie between 0 and 1, is called the trend smoothing factor. So, when we talked about EMA in the last lecture, you had 0 less than alpha less than 1, which was the weighting parameter. Similarly, 0 less than beta less than 1, or beta is called the trend smoothing factor here.

So, these are the underlying steps which are involved in the double exponential smoothing. Now, how do you extend that? So, can you extend this even further? Of course, we can, and the answer is triple exponential smoothing, or in other words, it is also called the Holt-Winters approach, Holt-Winters method, or Holt-Winters technique. Now, again, a few definition pointers or some general properties of triple exponential smoothing.

So, when do you apply that, or how exactly is triple exponential different from double exponential? So, we will discuss all that. So, before we discuss any of the points, just remember one very important idea: what happened in double exponential smoothing was that you still had the weighting parameter alpha, and then you controlled an extra idea, which was in terms of a linear trend in the model. But then, if you remember, we ignored seasonality completely. So, can you somehow bring in seasonality also?

And the answer is triple exponential smoothing. So, such a technique, which is called triple exponential smoothing or the Holt-Winters technique, is used for forecasting the time series when the data has both a linear trend and a seasonal pattern also. So, there should be some underlying linear trend, and there should be seasonality which is seen in the data also. So, hence, if you want to control both for the trend aspect and the seasonality aspect, we cannot use double exponential smoothing. We have to rely on something called triple exponential smoothing.

And this is also called the Holt-Winters technique, the Holt-Winters method, or third-order exponential smoothing. So, now do you see the progression? So, we started with SMA, where all the weights were equal, then we discussed EMA, where you can bring in the idea of weighting, right, and then we involved the alpha parameter. Then, the first thing in today's lecture is we discussed Holt's technique, where we kind of extended the EMA by bringing the trend idea also. And then, Holt-Winters is we are extending Holt's technique by bringing seasonality also.

OK, so if you want to draw a rough path, right, so it starts like this. So SMA and then EMA. So what is extra here? So the weighting idea or the weighting parameter alpha is extra here. Now, as you progress from EMA, then to the next thing, which is Holt's technique.

So, again what is extra here? So, the extra thing is the trend here, right. So, you are trying to capture the trend idea also, and then from Holt's, if you go on to Holt-Winters approach kind of an approach, then what is extra here? So, the extra idea is seasonality, right. So, you are basically adding one aspect to the earlier technique, which is a kind of a very standard extension or standard progression.

And then here I can write down plus plus everywhere. So, weighting plus trend plus seasonality is Holt-Winters. So, here if you want to control for both the trend idea and the seasonality, we want to introduce two terms here. So, as opposed to Holt's method where we introduced alpha and beta. So, here we want to introduce two extra terms to take care of the trend idea and the seasonality present in the time series.

So, one term would take care of the trend, and then the other term would take care of the seasonality or the underlying seasonal fluctuations. And such a triple exponential smoothing technique is capable of capturing an increase or decrease in the linear trend and the seasonal patterns. So, both trend and seasonality. So, now some involved notations, and then slowly we will try to build the model. So, ST is what?

So, ST is nothing but the general notation for the smooth statistics. So, we have been following ST in almost all the equations so far. Then the second one is alpha. So, alpha is what? So, alpha is a smoothing parameter of the data, and then alpha has to lie between 0 and 1.

So, the smoothing parameter is kind of equivalent to a weighing parameter. So, alpha is a weighing parameter. Then what is Bt? So, we saw in the double exponential situation. So, Bt is the best estimate of the underlying trend at a particular time t. And then we introduce beta also.

So, what is beta? So, beta is a trend smoothing factor, which also has to lie between 0 and 1, by the way. Now, here, when we talk about the Holt-Winters technique or triple exponential smoothing, we have to involve one more process, which is CT. Now, what is CT? So, CT is nothing but a sequence of seasonal correction factors at time t. So, BT stands for the estimate of trend, while CT stands for the estimate of seasonal fluctuations or a sequence of seasonal correction factors at time t.

And, similarly, we have to introduce one more parameter, which is gamma. So, gamma is the seasonal smoothing factor or the seasonal change smoothing factor. So, beta is my trend smoothing factor, gamma is my seasonal smoothing factor, and alpha is the weighting parameter or the general smoothing parameter of the data, right? And then all three, right? So, alpha, beta, and gamma have to be between 0 and 1.

So, these are some involved notations when it comes to kind of establishing that model of triple exponential smoothing. Alright, and then, further, let capital L denote the length of the cycle of the seasonal change. Now, since we are trying to bring in the seasonality idea also, then how do you capture the seasonal variations or of what length is each season, and so on and so forth, right? So, let capital L denote the length of the cycle of seasonal change. So, I will give you a small example: if you have monthly data having seasonality of period 12.

So, let us say temperature data or rainfall data, etc. So, if you have monthly data with seasonality of period 12, then my capital L would be nothing but 12. And similarly, let capital N denote the number of cycles. So, how many cycles are there in the entire data set? Now, again if you take the same example, if you have monthly data with seasonality of period 12 over let us say 10 years, then my capital L is 12 and my capital N is 10, because you are collecting data over 10 years.

So, you have to understand all the differences between the different notations. So, capital L stands for the length of the seasonal cycle. So, if you have monthly data, let us say temperature data, rainfall data, whatever, then my capital L would be 12. And then how much data have you collected? So, the number of cycles in the data set is given by capital N. And here, surprisingly, we have to capture two different cases of seasonality.

So, in the literature, one can actually identify two different cases of seasonality. The first one is called multiplicative, and the second one is called additive seasonality. So, it could be either multiplicative seasonality or additive seasonality. So, probably in the next slide, we will describe the difference between the multiplicative aspect of the seasonality and the additive aspect. Alright, so firstly, what do you mean by additive seasonality?

$$Y_t = T_t + S_t + e_t$$

So, obviously, I can actually write down the model structure in some kind of situation. So, YT equals TT plus ST plus ET. Alright, where obviously T is the trend, S is the seasonality, and E is the error. I think we discussed such a kind of model long back in some previous lectures. But again, this is just to tell you that the seasonality aspect has been added to the model, and hence it is called additive seasonality.

So, in other words, the seasonal effect is added to the trend. So, T plus S. Hence, the seasonal effect is roughly constant over time. So, since you are not multiplying it, since it is not proportional, you are simply adding the seasonal effect. Hence, the seasonal effect stays the same or roughly constant over time. Now, can you describe this using some example?

Of course. So, let us say we will take a small example here. So, imagine sales of a product over the year. So, you have some product, and then you have some sales data over the entire one year. Now, if you have an additive kind of model structure or additive seasonality.

The sales might increase by a fixed number, let us say 100 units every December due to some event. Let us say Christmas or New Year, whatever. So the sales might increase by a fixed number, let us say 100 units every December due to holiday shopping, regardless of the general sales level throughout the year. So, whenever December comes, you have a fixed number which is an increase in the sales. So, this December there will be an increase of 100 units.

Then the next December there will be an increase of 100 units. So, if you notice this number 100 stays constant over the entire time frame. So, if you collect data for let us say the next 10 years or so, then every December there will be a fixed increase which is by 100 units. And why exactly? Because you are adding the seasonal term to the trend.

Okay. And as opposed to that, what do you mean by multiplicative seasonality? So, under multiplicative seasonality, you want to multiply the seasonal effect by the corresponding trend. So, something like this. So, YT would be the trend.

$$Y_t = T_t \times S_t \times e_t$$

So, TT into ST into ET. So, T stands for trend, S stands for seasonality, and E stands for error. Now, again, the only difference here, as compared to the previous slide, is that instead of adding the seasonality aspect to the trend, you are basically multiplying it. Hence, the seasonal effect is multiplied by the trend, resulting in larger seasonal fluctuations when the time series is at a higher level. So, now can you see that?

So, let us say, since you are multiplying seasonality with trend, whenever the trend is high or whenever you have an increase in, let us say, production, sales, stock price, or whatever, right? So, whenever the time series is at a higher level, there will be a much more predominant influence of the seasonality, or there will be larger seasonal fluctuations simply because you are multiplying the seasonal effect with the trend. So, higher values would produce larger fluctuations, and a lower trend would produce lower seasonality as well. Now, again, the same example. So, imagine the sales of a product over the year.

But now, if you bring in a multiplicative model, the sales in December might double compared to the other months. Compared to the other months, or as we saw in the earlier example as well. So, in the earlier example, when we took additive seasonality, you had a constant increase every December. But now, since you are multiplying the seasonality with the trend, whenever December comes, the sales might double compared to the other months. Okay.

So, this is the difference between additive seasonality and multiplicative. And hence, we have to create different model structures if we want to account for proper seasonality. So, be it additive or multiplicative. Now, finally, coming to the model structures, right? So, again, the trivial situation is that S0 is Y0, right?

And by the way, all these are steps for a multiplicative kind of seasonality initially, okay? So, steps for multiplicative seasonality. So, S0 equals Y0. And what next? So, ST equals alpha into YT divided by C of T minus L. Now, why divided by?

$$s_0 = y_0$$

$$s_t = \alpha \frac{y_t}{c_{t-L}} + (1-\alpha)(s_{t-1} + b_{t-1})$$

$$b_t = \beta(s_t - s_{t-1}) + (1-\beta)b_{t-1}$$

$$c_t = \gamma \frac{y_t}{s_t} + (1-\gamma)c_{t-L}$$

Because you have a multiplicative effect. And plus 1 minus alpha applied on ST minus 1 plus BT minus 1. Now, in a way, the first equation is again similar to what we saw in the Holt technique or what we saw in the EMA technique. The only difference is what variables you bring in now. So, I think this thing is still the same.

But then, what do you multiply alpha with? So, instead of simply y t, I am multiplying alpha by y t divided by the seasonal factor. And again, why divided? Because you have a multiplicative seasonality. And again, on the other hand, how do you control for the trend aspect?

So, I think this is unchanged. So, Bt is still equal to beta into St minus St minus 1 plus the remainder, which is 1 minus beta applied on its recursive structure, which is Bt minus 1. And lastly, how do you control for seasonality? So, Ct is gamma into Yt divided by St again plus the remainder, which is 1 minus gamma applied on or multiplied with Ct minus L. So, in a way, this is the similar kind of structure that we saw in, let us say, either Holt's method or EMA.

Then, this is controlling for trend if you want to specify, and then the last equation controls for seasonality. So, this path is for a multiplicative kind of seasonality. Now, what would happen if you have an additive seasonality? We will discuss that in the next slide. So, if you have an additive seasonality, the things slightly change here and there.

$$s_0 = y_0$$

$$s_t = \alpha + (y_t - c_{t-L}) + (1 - \alpha)(s_{t-1} + b_{t-1})$$

$$b_t = \beta(s_t - s_{t-1}) + (1 - \beta)b_{t-1}$$

$$c_t = \gamma(y_t - s_{t-1} - b_{t-1}) + (1 - \gamma)c_{t-L}$$

Now, again, the trivial situation does not change. So, S0 equals Y0. Now, if you notice here, ST is alpha plus YT minus CT minus L plus 1 minus alpha still applied on ST minus 1 plus BT minus 1. So here, the only difference is since you are adding the seasonality, this is like YT minus CT minus L. So instead of YT divided by CT minus L, you want to subtract the seasonal effect from the series. So, YT minus CT minus L.

And similarly, I guess this does not change. So, BT is still beta into ST minus 1 plus 1 minus beta into BT minus 1. And lastly, you will have a similar-looking change in the seasonal equation. So, CT equals gamma into YT minus ST minus 1 minus BT minus 1 plus 1 minus gamma into CT minus L. So, depending on what kind of seasonality one has, either additive or multiplicative, the idea is how you remove the seasonal factors of the series.

So, either divide or subtract. So, if you have multiplicative seasonality, you want to divide it. If you have additive seasonality, you want to subtract that. Alright, so now, towards the end, we will discuss one numerical example and then we will see how you apply Holt's technique, how you apply Holt-Winters' technique, and essentially what difference you have, right. So, probably we will discuss this dataset long back in some of the earlier lectures also.

So, this is a simple-looking data on monthly air passengers: the number of passengers traveling by flights on a monthly basis over all these years. So, let us say starting from, I think, 1948 or something like that, all the way up to, let us say, 1970 or 1968, something like that. So, over all the years, right, this plot kind of tells you the behavior of, or the number of, air passengers on a monthly basis, right, and then this is the graph of the data. Now, again, immediately, whenever you analyze or you are presented with any graph in a time series course or a time series kind of exercise, immediately one should try to find out some underlying pattern in the data set. So, what do you see?

So, I can clearly see a trend in the data. So, since the values are increasing, you have an upward trend. Anything else? So, there could be some seasonality also. So, since you are seeing some fluctuations which are kind of even also,

Right. So, there could be some seasonality, which makes sense also, because the number of people traveling by air or by flights is seasonal. Right. Because in the summer, when you have more holidays or when you have holidays, then you'll have more number of passengers, maybe for vacation and so on and so forth. As opposed to, let us say, in between months.

Then again, probably in the winters, people would travel more because of holidays. And all these things. So air travel is supposed to be seasonal, which is being reflected here, is my idea. Anything else that you see? The answer is yes.

So, apart from that, let me write it down. So, apart from trend and seasonality, I can also see changing variance. So initially, if you see, the peaks are kind of subdued. So you do not have high peaks here initially. But as you go down the line, the peaks are much higher.

For example, here or here, as compared to let us say somewhere here. This means that my variance in the underlying data is also changing. So such an example contains all three aspects. So trend, seasonality, and changing variance. But my question is, can you try to model such data using some smoothing technique, which is let's say SMA or EMA or the Holtz method?

The answer is no, because you also have to capture the seasonality. So, Holt's technique would be able to capture, let us say, the general smoothing and the trend, but then one would require triple exponential smoothing or Holt-Winters technique to capture the seasonality aspect of the data also, and this is exactly what we will do. If you run a few codes in R, which we will see in the subsequent lecture, you will get an output which looks something like this. So, by the way, it is called Holt filtering on the data. So, what do you mean by filtering?

So, filtering is an alternative term for smoothing. So, if you apply a proper Holt technique or Holt smoothing technique on the data or double exponential smoothing technique on the data, it is also called filtering. So, filtering is nothing different from smoothing. So, what output do you get? So, if you apply Holt filtering on the data, so here it says Holt-Winters exponential smoothing with trend but without seasonal component.

So, can you see that? So, the data contains a trend but then says without seasonality, which means Holt method. So, R has a common package or a common function called Holt-Winters. It is combined. And then you have an option of including seasonality or excluding seasonality.

So, if you include seasonality, it becomes Holt-Winters. If you exclude seasonality, the name is still Holt-Winters. But essentially, what you are doing here is the Holt-Smith technique or double exponential, right? One more way of identifying is that if you look at this guy, so gamma equals F, so F stands for false, right? So, we are not taking any gamma here, and then gamma is a seasonality parameter that we saw in the last slide. So, if you do not have any gamma parameter, which means that you are only controlling for the trend aspect. Which also means that you are only applying the Holt technique and then not the Holt-Winters technique, okay?

So, in a way, the R output may be slightly confusing because the name still suggests Holt-Winters, but then probably you have to dig further or dig deeper as to what exactly is happening. So, if you are only controlling for the trend without the seasonal component, you are basically applying the Holt method, right? The other way to check is if gamma equals true or false. So, if gamma equals false, you do not have any seasonality there, okay. And by the way, once you apply the Holt technique, these are the estimated parameters that R gives you.

So, my alpha happens to be 1, and my beta happens to be 0.00321, whatever. And then gamma is false because we do not have seasonality. And by the way, this is exactly how the Holt smoothing is fitted on the data or the Holt filtering is fitted on the data. So, the black curve is the actual data, and then the red curve is the Holt smoothing technique applied on the data. And then here, you can clearly see that it's kind of doing a very good job and so on and so forth.

And then forecast from a Holt's method. So if you want to forecast down the line, once you fit the Holt's smoothing technique on the data, then how do the forecasts look? So this blue line gives you the forecast along with the standard errors or the standard regions along the forecast. And lastly, what would happen if you apply the Holt-Winters filtering on the data or Holt-Winters smoothing on the data? And then here, clearly, you can see that we are actually applying Holt-Winters and not Holt's.

Why exactly? Because it says that Holt-Winters exponential smoothing with trend and additive seasonal component. So R actually also tells you that is it Additive seasonality or multiplicative seasonality. So, here we both have trend as well as seasonality.

And then this is the call aspect, and then these are all the fitted parameters. So, my alpha is 0.3266, my beta is 0.0057, and my gamma is 0.8206. And then lastly, what would happen if you fit the Holt-Winters filtering on the data and And of course, the results should be better. So let us say, so again, the black curve is the actual data, and the red curve is the Holt-Winters being fitted on the data.

And then these are the forecasts. So here, you can see that the forecasts kind of replicate the seasonality aspect, which is exactly what we want, as opposed to Holt's technique. So again, if you remember, if you go back, if you only apply Holt's technique. You are getting forecasts which look like a straight line, right? So, the forecasts are not able to capture the seasonality aspect in the data.

It is only capturing the trend aspect, which we do not want here because we know that the data set is seasonal also, right? So, what? So, if you apply the Holt-Winters technique, then the forecasts are much better because they are able to capture the exact pattern of both the trend and seasonality also. So, this is the idea of applying, let us say, either Holt's technique or Holt-Winters approach and so on and so forth. So, probably the next class would be a practical session where we will try to combine all the ideas we have studied this week and then try to implement the ideas in R by taking some practical data.

Thank you.