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 23: Smoothing Techniques (SMA, EMA)

Hello all, welcome to this course on time series modeling and forecasting using R. Now, again, just to briefly tell you where we are at, so this week we are mainly focusing on the idea of forecasting. Right. Now, again, just to give you a very quick refresher. So forecasting kind of follows the model checking aspect and the diagnostic checking aspect of the model. So once you frame the appropriate time series model and then you kind of ensure that all the assumptions are being met or not, which is called diagnostic checking, then the experimenter or the analyst may proceed with forecasting.

And if you remember in the last couple of lectures, we focused on the different ideas of forecasting where we took one particular example, which was the ARMA model, and then tried to forecast down the line in the future using a technique called minimum MSE forecasting. And in the beginning of the last lecture, we also discussed the difference between, let us say, all these alternative terminologies, such as prediction, estimation, or forecasting. So, again, the essence of forecasting is, with a handful of historical data, once you frame an appropriate time series model and you have kind of ensured that all the assumptions are being met or not, then one has to kind of develop suitable forecasts down the line or, as we say, suitable forecasts in the future. Now, in this lecture, we will focus more on another slightly different idea of forecasting called smoothing techniques. And in this slide in front of you, you can see that one can actually work with all these different kinds of smoothing techniques, starting with, let us say, simple moving average smoothing, exponential smoothing, or the third one, let us say, double exponential smoothing, or in other terminology, it is also called the Holtz method.

And lastly, one can focus on something like triple exponential smoothing, which is also called the Holt-Winters kind of approach. Now, firstly, we will understand what exactly one means by smoothing and then how do you apply a smoothing technique to kind of forecast down the line. Now, firstly, why smoothing? And probably this slide would tell you the idea or a small definition of what you mean by the terminology smoothing. So, when practical data contains a lot of random variations or, in other words, we call the random variations noise also, right.

Smoothing techniques are helpful in reducing such random variations or noise. So, essentially the idea of smoothing, or smoothening as we say, is to smoothen out a time series. So smoothing techniques are helpful in smoothening out a time series. And then, as the name suggests, what does smoothen out mean? So I will give you a very quick example.

So let us say if your time series data is stationary and so on, then let us say if the stationary time series data follows some random pattern. Now obviously, if you want to forecast down the line, you may want to avoid all the random fluctuations in the data. And just to smoothen out all these random fluctuations, or to make them more uniform, or to reduce the number of random fluctuations or noise in the data, we have to apply an appropriate smoothing technique. So once applied, they reveal the underlying trend for better forecasting. Now imagine a situation where the data contains something like, or the path of the data is something like that.

So maybe this is again a hypothetical example, but let us say if you want to draw the practical graph of the data, then assume that it follows such a pattern. Now here, as one can clearly see, initially you have a lot of random fluctuations. Then in between, you do not have fluctuations here, and then again towards the end, you have a chunk of random fluctuations or noise. Now one has to apply an appropriate smoothing technique here to smoothen out the random fluctuations, or to avoid or reduce the amount of random fluctuations one sees in the data. So eventually, once you apply an appropriate smoothing technique, probably this graph may look like this, where the number of fluctuations is kind of less.

So, can you see the difference now? And then, this kind of graph is able to reveal the underlying trend much more predominantly. So, as compared to something like this sort of graph where you have lots and lots of random fluctuations or unwanted noise in the data, the idea of a smoothing technique is to smooth out all the random variations or random fluctuations so as to reveal the underlying trend or the underlying patterns in the data set. So, the last comment is that a smoothing technique can also reveal any other important patterns in the data or any important patterns in the observations. So, I think we will now start with all the kinds of smoothing techniques.

So, we will start with a very basic one, which is called a simple moving average smoothing, or in short, it is called SMA, okay? Now, as the name suggests, this is a very simple and common kind of smoothing technique, right? So, the short form is SMA, and then SMA stands for simple moving average, right? I'm pretty sure that if somebody is into trading or, let's say, if somebody is into investing and has some idea as to how the stock price of a company moves or what exactly candlestick charts are, and so on and so forth. So, I'm slightly deviating from the topic.

But, on the other hand, if you have some knowledge on the trading aspect or the investing aspect and then you spend some time on, let's say, looking at the graph of a particular stock pattern or, let's say, looking at all the candlesticks that there are, then you must also have some idea about, let's say, what would happen if you implement some moving average on the candlesticks, right? So, all the ideas of, let's say, implementing an appropriate moving average, let's say, a 21 moving average or a 50 moving average or a 100 moving average. So, applying a moving average on the underlying price chart or the price series is nothing but a smoothing technique, okay? Now, I think the definition would be kind of very easy. So, a smooth series is derived from the average of the last k elements of the series.

So, under this SMA or simple moving average kind of smoothing, how do you actually obtain the smooth series? So, the smooth series is derived by simply taking the average of the last k elements of the series. And here, this k stands for the order. So, the k stands for the number of items in the past over which you have to sort of average out and then take the average of the last k elements. And then, this is a very simple description.

$$S_t = \frac{Y_t + Y_{t-1} + \cdots + Y_{t-k}}{k}$$

So, S t is the smooth series, let us say. So, S t equals the average of the past k elements. So, if you are currently sitting at y t, which is the current observation. So, what would you obtain if you take the average of the past k observations? So, y t plus y t minus 1 plus y t minus 2 up to y t minus k divided by k. So, a simple-looking formula and hence the name simple moving average smoothing.

Now, immediately I can write down a few special cases or a few shorter versions of this S t. So, let us say what would happen if k is 3, right? So, if k is 3, this would be nothing but y t plus y t minus 1 plus y t minus 2 divided by something like 3 here. Or in other words, if you have let us say k equals 4, then my S t would be nothing but y t plus y t minus 1 plus y t minus 2 plus y t minus 3 up to y t minus 4 and then divided by 4. By the way, here I have missed one observation here. So, let us say this would extend to y t minus 3 because y t is my current observation and then you want to average over the last k elements of the series essentially.

So, this is one way, or what you can do is you can take t minus k. So, if you have to stop here, so t minus 2, this would be t minus 3 and then plus 1. So, let us say t minus k plus 1, something like that. So, this is the idea of simple moving average smoothing, basically. Now here, we will very briefly discuss a practical example, and in this practical example is Google's stock price.

Now again, a short while back, we discussed that if you have any information about how the candlesticks behave or how the price behaves of a particular stock, then you must have followed such a graph in your lives. So, let us say the underlying shows the price pattern of the Google stock. And then on top of that, we have tried to fit all these different SMAs or simple moving averages on the underlying Google stock price. And then each color kind of contains some importance in a way that each color denotes different smoothing windows. For example, the blue color, we are averaging over the last 10 elements only.

The red color, we are averaging over the last 20 elements only. And if you look at the last one, which is given by yellow, we are averaging over the last 200 elements and then proposing the smooth series value. And again, inside this graph, you can actually see different curves having different colors. For example, the green one here or the yellow one is right there. Now, just for a second, if you pause the video and then just kind of digest this entire idea about implementing some moving averages on an underlying stock price, then out of all these different colors or all these different curves, which curve do you think is smoother as compared to the other colors or as compared to the other orders?

So, the answer to this question would be yellow, right? Because if you look at the yellow curve, where the order is the maximum, which is 200, by the way. So, the yellow curve is all this curve that you see here, which is much smoother than something like the blue curve. So, the blue curve you cannot even see here because the blue curve is so matching with the data.

In fact, the blue curve one can actually visualize here, for example, or here. So, can you see that the blue curve is almost on top of the price and then moves exactly as the price does? And why is that? Because the averaging window or the window size is small there, which is just 10. So, what happens is you are basically proposing one value for every 10 past values, as opposed to something like the yellow curve, where you are proposing one value for the last 200 observations.

So, if you are proposing one value for the last 200 observations, do you not think that you will have a smoother curve as compared to, let us say, something like the order being 10? So, this is the idea. So, depending on what the order is of the moving average, which is applied on the data, you will either get a less smooth curve or a smoother curve. So, now I think we will discuss another kind of smoothing technique, which is slightly different from SMA, which is called exponential smoothing or, in short, EMA. So, the short form is EMA, and EMA stands for exponential moving average.

So, again, it sort of relies on the idea of moving averages. So, earlier, we discussed SMA. So, SMA stands for simple moving average, and on the other hand, now we will talk about EMA, which is called the exponential smoothing technique or exponential moving average. So, what exactly is the difference between EMA as compared to SMA? So, we will find out now.

On one hand, SMA smoothing uses equal weights for all the observations. So, again, just for a second, if you go back to this formula here, then you do not have different weights for any of the observations. So, this is a very, very simple-looking average or standard average of the last k values. So, the SMA technique does not assume any differentiating weights on the observations, whereas if you talk about something like EMA, the exponential moving average or the exponential smoothing technique uses exponentially decreasing weights over time. So, exponential moving average or exponential smoothing could be kind of compared to something like a weighted average instead.

So, rather than taking all the weights to be equal applied to all the observations, what would happen if you take some exponentially decreasing weights and then apply them to the observations? Okay. Now, here, one very important point to make is that probably I will ask all of you a very simple question: let us say if you want to apply the weights on the past observations, right? So, let us say you have yt here, and then you want to apply some weights on yt-1, yt-2, yt-3, etc. Now, just for the sake of applying the weights, on which observation do you think that you should put more weight?

So, would that be something like yt-1, or would that be something like down the history, something like yt-10 or yt-15, whatever? But my question is, on which observation would you put slightly more weight? Now, again, the obvious answer or the common-sense answer would be the very recent history or the very recent past, which is yt-1. And this is exactly what we will do. So, we apply more weight here on the very recent history because we kind of assume that history would repeat itself.

So, something which happened last month or something which happened, let us say, 2 months back. I should give more weight to those observations as compared to something which happened, let us say, 10 months ago or 15 months ago or 20 months ago and so on. I will give you one very easy example in this regard. So, let us say you are dealing with monthly temperatures, and then you have data of all the monthly temperatures over, let us say, 5 years or 10 years. Now, let us say you want to apply the exponential smoothing technique or the EMA, and then you want to apply it to the temperature data.

Now, let us say currently you are sitting in October, and then my question is, on which month or to which month's temperature data should you apply the maximum weight? Now, obviously, the answer would be September, right? Because if you want to forecast down the line, let us say November month's temperature or December month's temperature, etc. Then one would want to put more weight on the recent history rather than something like, let us say, January temperature or February temperature or March temperature. So, hopefully, the idea is clear that we have to apply exponentially decreasing weight. So, let us say we apply more weight on this guy, then slightly lesser weight on this, then even lesser weight on this, etc.

So, in this period, we can write down the model in a recursive manner for EMA. So, what happens is, for starting the model, we require some starting point. So, S0 would be Y0, let us say initially, and then for any other t, my s t or the smooth series would be nothing but some alpha into y t plus 1 minus alpha into s t minus 1. So, now clearly, can you see that this is a recursive structure in terms of s t? And then here, the alpha is nothing but called a smoothing factor. So, alpha has to be some weight which has to lie between 0 and 1.

$$S_0 = Y_0$$

$$S_t = \alpha Y_t + (1 - \alpha) S_{t-1}$$

Where, $0 < \alpha < 1$ is the smoothing factor.

So, alpha has to lie between 0 and 1, and alpha is called a smoothing factor. Now, immediately, a couple of questions would be asked here. So, what do you think would happen if my alpha is really large, firstly? So, what do you think? So, you can pause the video and then try to answer this question: if I bring my alpha or if I go on increasing my alpha, in that case, what would happen?

So, if I go on increasing my alpha, I am putting more focus on the recent observation, which is yt, rather than on the past, which is exactly what we want, right? As opposed to what would happen if my alpha is really small. So, if my alpha is really small, I am not putting enough weight on the recent observation, and I am putting all the weight on prior observations, which is against the assumption and could be a bad signal for me. So, in this sense, I have the liberty of choosing or controlling the weighting factor, which is alpha. So, I can put alpha to be, let us say, 0.9, or I can put alpha to be 0.1, or somewhere in the middle, let us say 0.5 or 0.55 or 0.6 or 0.4, something like that.

So, EMA is kind of advantageous compared to SMA in the sense that I have control over how to assign weights to all the past observations, right? Alright, so probably again the same kind of graph of Google stock price, but instead of applying SMA, we are trying to apply appropriate EMAs. And then here, instead of orders, as you saw in the SMA graph, we have the liberty of choosing alphas now. So, probably the blue curve gives you alpha with 0.01, then the red curve is alpha with 0.02, then the green curve is alpha with 0.1, the purple curve is alpha with 0.2, and lastly, the yellow curve has the maximum alpha, which is 0.99. And again, the same kind of situation.

So, you have the underlying Google stock price over those many years, and then each color represents a different EMA fitted on the data having different alphas. Now, here are a couple of observations immediately. So, if you compare, let us say, the blue curve and the yellow curve. So, what characteristics does the blue curve have? So, alpha is the least.

So, alpha is 0.01, and then clearly, can you see that if alpha is 0.01, you are basically giving more weightage to earlier observations rather than the current ones or the very recent ones? So, my EMA should be very close to the actual price as opposed to something like alpha being 0.99, right? If I have a very large alpha, that would give me a much smoother kind of curve, okay? So, in a way, I do not want a very large alpha, and obviously, I do not want a very small alpha. So, I should strike a balance between the values of alpha, all right?

So, probably what we just talked about is kind of given in this slide here. So, a lower alpha means lower influence of each observation on the smooth series, right, and hence the EMA would be less responsive to recent changes. So, whatever changes you kind of reflect in the recent history, the EMA would be very, very less responsive to all those changes because my alpha is kind of very small, right. So, if my alpha is small, it means that I am putting lower influence on each observation on the smooth series. On the other hand, if my alpha is large, a larger alpha reduces the effect of smoothing and gives more weight to recent changes in the data.

So, if my alpha is very large, something like 0.9, 0.99, or 0.85, etc. So, a larger alpha reduces the effect of smoothing and gives more weight to recent changes in the data. Thus, alpha being 0.99 shows a much smoother curve as compared to alpha being something like 0.01 or 0.02, etc. Okay. So, this is broadly the idea of EMA and then what exactly is the difference between EMA and SMA.

So, again, just to summarize, in EMA, we are putting some weights to all the observations as opposed to equal weights being put in SMA. Okay. So, probably one last slide on EMA that you have some properties or some use cases of EMA. So, in this case, what happens is that the forecast is constructed using an exponentially weighted average of past observations. So, instead of a simple moving average, we are again talking about an exponentially weighted average of all the past observations.

The second point is that more recent values have a greater influence on the forecast, which is exactly in line with the assumption we have that recent history tends to repeat itself in the nearer future rather than the past. And there is an exponential decay of the influence of past data. So, the most recent data point is the most influential. Then, the influence decays exponentially. Yet, the EMA technique is simple, computationally efficient, and allows for ease of adjusting to changes in the process being forecast with reasonable accuracy.

So, here, one point to make is how one can adjust to changes in the process by applying different weights immediately. So, rather than applying, let us say, a weight of 0.8 initially, if I see that the data is shifting or the data has some structural changes in the nearer future, I can change the weighting mechanism so that the recent observation now has a weight of, let us say, 0.7 or 0.9, which is different from the earlier 0.8. And the last point is that it allows one to determine the influence of recent observations on the forecasted values. Now, what exactly is the model aspect of the underlying EMA? So, now, since EMA is also used in forecasting, one can derive the following equations.

You have to assume these two characteristics. So, let us say ST equals YT plus 1 hat. Now, remember from the last lectures, what do you mean by YT plus 1 hat? So, YT plus 1 hat is the one-step-ahead forecast. Similarly, my ST minus 1 would be YT hat.

$$S_t = \hat{Y}_{t+1}, \; S_{t-1} = \hat{Y}_t$$

Thus, can I rewrite the earlier equation that we had? So, if you go back a couple of slides, this was the initial EMA equation that we had. So, ST equals alpha YT plus 1 minus alpha ST minus 1. So, can you sort of replicate the ST and ST minus 1 in terms of some hat values? Of course, and then this is the answer.

$$\hat{Y}_{t+1} = \alpha Y_t + (1-\alpha)\hat{Y}_t$$

So, instead of ST, I will write down YT plus 1 hat, and then instead of ST minus 1, I will write down YT hat. So, the new equation becomes YT plus 1 hat equals alpha YT plus 1 minus alpha into YT hat. Now, what? So, one can actually apply a recursive kind of situation now. So, again by recursive substitution, if you start here, let us say, then again I can replace YT hat with its recursive value, which is YT minus 1 hat.

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t$$

$$= \alpha y_t + (1-\alpha)[\alpha y_{t-1} + (1-\alpha)\hat{y}_{t-1}]$$

$$= \alpha y_t + (1-\alpha)\alpha y_{t-1} + (1-\alpha)^2\hat{y}_{t-1}$$

$$= \cdots$$

$$= \alpha y_t + \alpha(1-\alpha)y_{t-1} + \cdots + \alpha(1-\alpha)^{t-1}y_t = \alpha\sum_{j=1}^{t-1}(1-\alpha)^j y_{t-j}$$

Then here again I can replace YT minus 1 hat by its recursive value and so on. So, this is equal to alpha YT plus 1 minus alpha into alpha YT minus 1 plus 1 minus alpha YT minus 1 hat. Now, this entire thing that you see inside the square brackets is nothing but a recursion of YT hat using the starting formula. Can you see that? Because since YT plus 1 hat takes that form, I can go on replacing a similar structure in terms of all the hat values.

Now, again here, I can replace a similar structure in terms of y t minus 1 hat to reduce to that, etc. So, eventually, what will I have? So, eventually, I will have this recursive sum, which could be concisely written down in this form. So, alpha into summation 1 minus alpha to the j y of t minus j, j going from 1 to t minus 1. And then, the expansion of this summation is given by this kind of an equation here.

So, this is the underlying module structure that goes inside the EMA once you apply the appropriate weights, which is given by alpha. Now, probably here, we will spend some time on the question of how to select the appropriate alpha, right? I mean, now that we are discussing that the EMA kind of focuses on a particular value of alpha, right? And then, we discuss what exactly do you mean by alpha? So, firstly, what is alpha?

So, alpha is nothing but the weighting parameter, right? So, if you have a larger alpha, we are giving more weightage to the recent history, and if the alpha is kind of very small, let us say 0.01 or 0.02 or 0.1 or something like that, we are not giving more weightage to the recent history. So, now the question is, how do you select the appropriate alpha or how do you select the alpha for a particular problem? So, in this case, MSE or RMSE can be used to select the value of alpha. Now, what exactly do you mean by MSE?

So, MSE is mean squared error. And then, what is RMSE? So, RMSE is root mean squared error. Now, of course, these two are kind of analogous. So, one can either use MSE or RMSE to select the value of alpha.

But how exactly? So, we can do an empirical kind of study or a simulation kind of approach. So, what we will do is we will assign values of alpha ranging from 0.1 to 0.99. So, let us say 0.1, then 0.11, then 0.12, then 0.13, etc., all the way up to 0.99. Because we know a fact about alpha that alpha should lie between 0 and 1.

So, what we will do is we will take some equidistant values of alpha between 0 and 1, starting with 0.1 all the way up to 0.99, and simply try to model using all these different kinds of alpha values. So, for a particular alpha value, I will propose a smoothing technique, right? So, in a way, how many smoothing techniques would you have? So, you have that many smoothing techniques. So, 0.11, then probably the next value of alpha could be 0.2, then the next one could be 0.3, 0.4, 0.5, let us say, all the way up to 0.99.

And once you apply all these different smoothing techniques having different alphas, select the value of alpha with the smallest MSE or the smallest RMSE, right? So, here essentially you will have 99 such smoothing curves applied on the actual data, right? So, out of all those 99 smoothing curves, pick the curve where you get the smallest MSE or pick the value of alpha and the corresponding curve having the smallest MSE or the smallest RMSE, right? Now, just for an example, so the example that we took of applying some EMA on Google stock price, right? So, if you code this in R and then do the entire exercise, it so turns out that the optimum value of alpha comes out to be something like 0.67, okay?

And of course, we will see in more detail during the R session, but then what exactly does this value mean? It means that out of all these different fitted EMAs on the Google stock price, the EMA curve or the EMA smoothing technique having this alpha value of 0.67 produces the least MSE or the least RMSE. And probably in the last lecture or a couple of lectures back, we studied that MSE or RMSE could be a criterion for comparing the forecast or a criterion for comparing the model fit. So, this is one point to remember: how do you select alpha. So, alpha should not be kind of put randomly, right? I mean, we have seen the drawbacks of that.

So, if you pick alpha to be, let us say, really high, 0.99, you also have a drawback. If you put alpha to be very small, it is a drawback in its own sense, right? So, one should always take a balance and then, depending on the practical data or the nature of the data, one should select the alpha rather than proposing alpha and then trying to fit the model using that alpha only, okay? Alright, so probably now in the next lecture, we will kind of extend the idea of smoothing techniques to, let us say, double exponential and then triple exponential, etc. Thank you.

Thank you.