ADS-506: Applied Time Series Analysis

Lab 2.1: Smoothing Methods

Hi everyone, welcome back. In this video we are going to cover exponential smoothing, which is also rolling averages or weighted averages, and we're also going to cover the ETS function. So, yeah, it'll be fun.

All right, so first thing's first, let's load in our library. So, I'm going load in the zoo, that's going to help us with our rolling averages. And then I'm going to load in reader. All right, cool. Let's set the seed, 406. All right, so let's load in the data set. I'm just going to import it and I'm going to load in the struggling wines data set. Cool, looks good. Not going to make any changes here.

All right, so this is loaded. And if we want to view it it's right here. I'm going to work with the reds, so let's just filter out what we need. So, we have reds. And I'm going to convert this to a time series object. So, struggling wines, right? Reds. Now, it starts at some point, so when does it start? So, our first set was actually January of 1980. So, we can see start. And the year is 1980. And we're going to say one for the first month, which is January, and then frequency 12. So, by default it has seasons. And let's plot this.

All right, so let's load this up and plot it at the same time. As we can see here, reds is now my time series object. It exists in my environment, so I can call it whenever I want. And then here's the plot. So, we can see there is an upward trend and there's definitely seasonality to this. Every now and then you might get a data set and you might not see the seasonality or the trend so easily. And you can use the ACS plot, the auto-correlation function, and just run this. And so that tells us how each lag correlates to the previous lag or to the first lag. So, this is where we're starting and we see here one. This just means first season. This is actually going to be the twelfth lag, here's the 18th. So, this looks about 22 lags. We can change it, so we can view more lags.

But really what we're looking at is there is some seasonality. If we're only seeing one season it might be kind of hard, so we can see log [inaudible 00:04:26] let's look at four season. So, that's going to be 48. And so we see every season it does have a peak and it correlates to the previous season. And we also definitely see a downward trend.

The other thing of note is we see all of these correlations are statistically significant because they are above the blue dash lines. So, when making our models we're definitely going to have to deal with this. I would say half of the models don't like seasonality or trend. So, we need to de-trend it and de-season it. Other models, they don't care, right? So, if we're working with neural networks they don't care. We can just leave it alone. But if we're dealing with exponential smoothing, things like that, then we do need to take care of this. And that's part of the pre-processing, right? Just looking at the data and see, okay, what's it telling us? What do I have to do to use this model? And so that's just part of it, just looking at it. And so we're going to de-trend this and de-season this. But for now we can just take a look at the moving averages.

So, we would look at moving averages to see the general direction of where the plot is going, right? So, we use this function called roll mean, and it is from the zoo package. So, if we just do question mark and in the function, then brings up that help menu and then we can see all the prior orders that go into it and what their definitions are. So, we're going to be dealing with X and K and a line.

So, X is going to be the object, right? So, that's what we're going to be working with. So, let's create a new object. So, we'll just call it moving average. Moving average. And this will be roll mean. Dealing with reds. And so K, that is the number of lags we're going to look into, right? So, if I set it at four, it's going to look at time right now, the time before, the time before that and the time before that. So, we're going to get four values and average them out. And then that'll give us the mean for those four and so on.

And we need to align it. And so we're going to align right. Now, we could align center, and what that'll do is if we're talking about T being our time right now, then it would be T minus one and T plus one. So, we're going to look back and forth. But if we were dealing with financial data, that's not really helpful because we don't know what's going to happen in the future. If we did we wouldn't be working. We would be really rich. So, we're going to look at align right, so that's just going to be past values.

We can just run that and we see here that it's actually leaving out the first three, and that's because we told it to look back four values. And so if we started here, there's not four values back, so it gave us a null value. Same thing with February, same thing with March. April is the first month that we can actually back and get four whole values, which would be January through April. So, this would be the rolling average, and I'll show you right now how it smooths out.

[inaudible 00:09:22] series, actual [whole air 00:09:30]. Run that. Okay. So, we see here actual is going to be the red, and then the moving average is this blue. And we can see it is smoothing out our data. And so now we definitely see that sine curve and we definitely see it going up.

And, like I said, moving averages are done quite a bit in finance. Especially if we're dealing with stock prices. So, a lot of times though what we would want to do is a weighted average. So, this example that we did is just a moving average where each lag or each month before it has equal weight. So, basically what we're saying is April is equally affected by what happened in January, February, and March, right? In terms of finance and stock prices that is not the case. More recent values should be weighted more and so that's exponential smoothing, right? And so let me show you an example of that.

So, we're going to load in this package called quant mod. And what's that going to do is that gives us another set of functions where we can get ticker prices. All right, so get symbols. And so this is getting prices for whatever ticker symbol we put in, right? So, let's pick … Anyone? Anyone care? Speak up? No, all right. Let's pick Apple then. I'm just kidding. There's no one here, it's just me.

All right, let's just say we are going to pick 2015. That was a good year I think. I honestly don't know. All right, so let's run that. I ran it and we now have another object, time series object, Apple. And so if we just want to see this, Apple gives us a few columns … It's actually not just Apple. Get symbols gives us a few columns for this stock. We get the opening price, the high for that day, the low for that day, the closing price, the number of shares traded and then the adjusted price per share. So, let's just dealing with the closing price. So, I'm just going to call it Apple, I'm going to spell it correctly first. I'm just going to get the closing price. Again, if I just run this we see the price, it's just the closing price.

All right, okay, so now we're going to deal with exponential smoothing. So, I'm actually going to give the most recent value heavier emphasis. So, before we do that, let's create our training set and testing set. So, we have our testing and so I'm going to say window. Because we're dealing with time series we can use the window function. So, Apple. And our training set, I'm going to end it at the 241st value.

If I had dates I could just say, "End it on December 15," right? So, we can just say this will end on the … No, I do have dates. Sorry about that. So, 12/15. So, what'll happen is it ends, it gives me everything all the way to December 15.

All right, so Apple test because that's what we're going to validate with. And we now have 2015. So, the training set ended on the 15th, the test set will start on the 16th.

Okay, and so if we wanted to make a model, we're going to make a model using the training set. So, let's make our model. And we are going to use the exponential smoothing, which is SES. Apple train. Right? Now, I'm going to set the alpha. What this alpha does is it tells me how much weight to give the first value, the first lag, and then it tapers off exponentially after that until we get to zero. And I'll show you why I picked or how you can pick the alpha. So, alpha, 0.8, level. So, there's our model. We want to see it. So, Apple predict. So, I'm going to tell it … I made this forecast so many moments into the future, so let me see what the length of my test set is. 10. All right, so we're going to forecast using our model, and H is equal to 10.

All right, so now if we wanted to plot this, we go auto-plot and I'm just going to pick Apple. And [inaudible 00:18:54] layer, that's going to be our predictive values. Predicted. So, I'm going to set this alpha now. This alpha is different than the one from our model. This alpha refers to transparency in our graphical objects. So, I'll just say 0.4, just so we can see through it.

Okay, so looks like it didn't like the dates after all. All right, let's change this and that would be … End would be 241, would be 251. And we are going to make this a time series. I'm not specifying dates, start time, end time or frequency, just a basic … Actually this is 42. There we go. All right, that's a little hard to read. So, let's zoom in. So, this will be X limit. And let's look at two [inaudible 00:20:56] one.

All right, so what I did was I just zoomed in just so we can see. Now, we gave it … Here's our predicted value. So, this is where we're expecting the prices to be. And so this cone shape, that is going to be our 95% confidence interval. And we can see that the reason it's flatlined here is because we don't have new values to give it. So, that's why it just flatlined because it just stopped at the last value. But we can still look at this model and see how well is it, and we can see the price. It stayed within our confidence interval, so it was below expected values.

And what we want to do is use this to make our predictions, but this really just tells us the nature of a plot and where it's headed. And then kind of the shapes of what is affecting it. So, if we wanted to see everything then we can just plot the fitted values and this is what the model looks like. So, we have our training set, which is the actual, that's going to be the blue line, and then we have our smooth model that's weighted. And so we can see it's actually pretty close. It does mimic the behavior of the actual pretty well.

Now, I picked the alpha. I set it when I created this model. I set it at 0.8. And so you might be asking yourself, "Well, how do I pick the alpha?" And honestly you would pick the alpha that has the lowest RMSE, the lowest root mean score error. So, what you would do is you're going to choose alpha with the lowest … All right, so if you're wondering how to find the RMSE, it actually doesn't even have to be the RMSE. It could be the mean score error, the MAPE, whatever metric you're using to define your models. I picked the RMSE.

So, you can say summary, which is a great function to know, Apple model. And so this summary gives you your forecasted values, but also gives you your error measures. And so you have RMSE there. If you wanted to see other models but you didn't want to plot everything, you can always just say summary. All right, so in here we have Apple train. And let's set our alpha at 0.5, right?

All right, so this value, our original one is at 0.8. And so this will be 0.5. All right, so we have 0.5, and now the RMSE to that actually went up to 0.54. So, actually has a higher RMSE. So, 0.8 is actually a better alpha. If we wanted to … Let's just change it, think okay, well, maybe higher is better. So, let's try alpha equals 90. So, what that means is yesterday's closing price has a huge impact on today's closing price. Today's three days ago closing price has a very minimal effect, and then past value you probably don't have anything. So, if we do this we have 0.9, and that one is … All right, so that one is 0.4947. So, that one is lower than the 0.8, but it's not really all that much and we can see that here.

So, if the 0.8 is our base model, then that RMSE minus the base or the old. And so it actually improved by 1%. So, this is where you have to take into consideration the amount of work involved in changing your models, even by just changing the alpha level from a 0.8 to 0.9. Is it really worth it? Are you really gaining a significant amount of accuracy for a decrease in error of 1%? Probably not all that much. But if you originally started with 0.5, and then we went to 0.8, right? So, that was an 8% decrease in our error. So, that's already better.

So, it just depends on what you're dealing with and what is a good value for you. You have to offset the amount of work involved in changing your models to what value is gained, right?

All right, that is exponential smoothing and moving averages. In the next video I'm going to cover differencing, first order of differencing, and how to de-trend or de-season a data set. Thank you.