ADS-506: Applied Time Series Analysis

Presentation 3.1: Autoregressive Models, Random Walk Explained, ARMA, and ARIMA Models

So the next model they have in the book, we'll start with this formula. The X at time T is some number times the previous value of this. So this is actually, the output is now being used, your old output is being used. Some other coefficient times your output from two samples ago. So T's the current time and T minus one, you can read it's like this, it's a subscript, right? So it's just X now, X one sample ago, X two samples ago, plus this noise. See how this is still the same T? So this model is called an auto regressive model. It's called a regression model because this should look like a linear a regression model, which we should be familiar with. And if not, again please review it in appendix B. But our regression model, this looks a lot like a linear regression formula. You have something here times coefficient, something else times coefficient, something else times this coefficient is one, right? So it's a linear model.

The auto part is, it's being predicted by itself. So where it's going to go next is based on where it was a little bit ago and then you're us subtracting where it was two samples ago. This WT, again, is a noise. It's the same WT that we had in that white noise model. So this equation, what it's saying is you look where you were and then you're adding some fresh noise and that's going to carry you forward to do the next thing. Then you can imagine how again you could write kind of a code like this in a for loop so you can start somewhere and run it a forward and keep pushing fresh noise through the model and it's going to look like something over here on the right. You can see that this auto regressive model is again less jagged or more smooth than the previous ones. So getting back to how we talk about how these models differ, this one definitely has some structure.

This, just by saying it's an auto regressive model, you're saying that the future depends on the past in a real strong way and how far back you look is an important thing to set in this kind of model. We look at this plot, it's the smoothest time series we've seen so far. So when we go to get into modeling the data, this is kind of the flavor of it.

Each term. So now we're saying that again our XT, our future is going to depend on adding that constant, adding the old value of XT minus one. So this is the past. This makes it look like an auto regressive model plus some fresh noise. So that our random walk with a drift is plotted in the bottom panel here. You can see that where you end up at Xt, you're adding these up over time and then you also adding some constant, then you have this little bit that depends on where it was. A walk you can think of as literally a random walk, or it's a lot of models of how particles bounce around a room. Your position isn't going to instantaneously jump because it depends on where you are. You can't teleport. But the air in this room is actually doing something like this. The molecules are kind of bouncing around and drifting and so they depend on their past, they depend on the noise. You can imagine I can inject some constant, like heating the room up and I'm going to be able to push them in a certain way.

Or maybe there's a way to compel them to all move to that side of the room, put a constant of force in. That's kind of these random walk models. These will also look like certain kinds of human behavior like when we look at the financial time series, you can see it's depending on the past some noise and some constant being added in. Just if you keep pumping money in, the price is going to keep going up. Then this drift is this kind of a constant thing on the graph, it's that dotted line. A drift is just sort of where you're going and then you're moving up and down from that drift. The drift is going to come back and be pretty important in the next chapter when we talk about some of the properties of these models, how they differ in terms what kind of data you're going to see. So the last example that they go through in the book, and this is actually my own work, the most common assumption I make is that there's a signal plus some noise.

So the noise is going to be, if you go back through everything that I've written here, you go through all the models in the first chapter there. The noise depends on a time and we're just adding in some, but the noise is the same. We're always adding noise because that's some of the randomness. The output of this is that second row in the plot where you get a sign wave and some stuff riding on top. That middle plot is the sum of that bottom plot, it's a noise. The top part is some deterministic signal. They give a thing like two cosign, two Pi, T plus one, five over 50. All right, this is a T, this is a plus. But this signal this is just giving you that cosign wave and it's just giving you some random noise.

Okay. So this first model, this linear regression model I've written up here, and you should remember from this one that your output of time, T, has these weights that are adding up what's coming in plus this noise term. So this is white noise as usual. These Zs are regressors. And these are also in the chapter called exogenous variables. The important thing is these Zs are actually coming from outside of the model. They're some other information. So if this is trying to predict, say, the surface temperature of the ocean, this might be the air temperature or the past weather or some other variables that aren't X itself. So the difference we're going to make in this chapter is we're going to replace these Zs with past values of X to give us this auto regressive model. In the same way we've gone from correlation to auto a correlation, we're going to do a similar thing from taking our basic regression model and doing a more time series aversion where you're using the past to predict the future.

Okay. So I've filled in the auto regressive model and what you should see is where we had these Zs in the regression, we're going to have past values of X. But remember X is the output itself. So it's not the same as in the regression chapter where you can do a lag, you can take Zs and look back into their past. This is taking our signal that we're trying to predict, like that surface temperature of the ocean and say something like the temperature it's going to be in an hour is the temperature it was an hour ago plus something of how the temperature was two hours ago and further back in time. So using this past information, which is how a lot of things work. You might imagine that the temperature is kind of smooth and might be oscillating as the sun comes up and down or similar, you could predict tides this way. It's a powerful but different model than the regression model.

These Greek letters here, these are fees, correspond directly to these regression coefficients. So they're telling us how much to add in of this part of the past and how much to add in from a further back. One of the big differences with this chapter and what a lot of this stuff is focusing on is the noise. So in here we have a white noise term, but you might also be having a moving average of your white noise, that's where the moving average part comes. We'll see that makes some important differences to the model. So we're going to skip to right now just the most vanilla simple auto regressive model. So I'm going to take this, I'm going to erase most of it and we're going to start with the simplest case.

So this is our first model. This is the AR1 model. So this is the auto regressive model that looks back one step in the past, adds that in, plus adding a new noise in at each term in the cycle. One of the key things about how these auto regressive models start to differ from regression models, in a regression model that beta coefficient can be any number that you want. Here, this fee has some constraints. If this number is greater than one, then the magnitude of the signal just grows. So this model is, over time things just increase more and more. One way you can get that kind of model is if you have a feedback. But the main thing is this will also blow it up and mean that you're not stationary, because of course your mean's not constant if you're always going up. So that's the first lesson is these can't be any number but these have to be a constrain to be less than equal to one.

This next way of writing is important and they're going to come back to it multiple times in the text, which is if you think about what this is doing or if you plug this model for itself and you go all the way back, you can see that ... Like see if you have this X of T, and then this is itself X of T equals fee times. Then we can plug this thing back in and get this T minus two plus some other weight and so on. You'll see if you keep substituting it and do all the algebra that you end up with a sum from zero to infinity of this fee times the old values of the weight.

So this is saying two things. One, it's saying that what happens next depends only on what the noise was and you're adding it up over time. Again this is just what happens from the recursion. Right? This line is saying you take where you are and look in the past and you see you have some version of the past scale plus the weight. But if you look two steps back in the past, you have some version of the past plus the weight. So you're really just kind of adding up these weights which gives you a different representation. It also gives you to this notion of the process being causal. This is very important, it's that these models are designed and restricted so they make sense, so that you can only predict the future from the past defense. If I know, this is not only useful but it prevents you from having a model that is implausible and uses information from the future to predict the path.

So if we go back, we have our auto regressive model. Then we also have our white noise model, which is just this noise term. If you think about what this white noise means, it means that at any time you're just taking a new sample, you don't really care what time it is, there's no relationship between adjacent time points. But this is different now because now for two time points that are close by, they're definitely related, you know how they're related by this term. So now times three where this fits is going to be one where the each individual time point is no longer uncorrelated like the white noise. So the next dimension to push into is a little bit about what that means for the errors to be correlated or uncorrelated like in this white noise model. So we're going to move on this moving average model.

So this is our moving average model and the thing here is in this auto regressive model, you're taking your past value and this new noise. All these WTs are always white noise. So you can have a model that's just white noise. We have this auto regressive model where you're adding new white noise at each step. Then we have this moving average model which at first glance looks kind of different because you have a bunch of different noise terms. So it's not just one white noise but it's a sum of multiple different white noises lagged in time with these new coefficients. One of the things that's that is worth looking at and the reason the book stresses all these mathematical definitions is now you can read off from this without having to test that you're doing any code from the equation. You can see that this matches that definition of a causal model because this XT is coming from the past and you can collect terms and write this in exactly the same equation we just had. It's also going to be a stationary.

Unlike here where we had some restriction that these fees got to be less than equal to one, there's no restriction on the parameters for the moving average model. We've seen a moving average model a little bit informally in the first lecture and you can think about it as having this window. So you might have the past couple of samples and then it drops off. It's like you're kind of just moving this window across your time series and just grabbing a few points at a time and averaging them. So the past when you get out of that window goes to zero. So this is one of the differences between this moving average model, which is that your auto correlation is going to drop to zero once you're outside, how far back you go. So this is the average of three terms. So if this ended that weight at time T minus three is not in this equation. You don't have to calculate it, you can just check. This thing can only depend on these three terms because the other ones aren't even in the model.

But this thing, this auto regressive model because it's defined in this recursive way, you have to go all the way back. That's one reason why we were writing it is that infinite sum. Overall the past values of the way you add them up with this term to give you the auto regressive model. One of the things that they point out, and this is different from the linear regression case. The way that the linear regression estimator is defined, you'll get a unique best regression fit. But for this moving average you can get multiple settings of these coefficients that'll give you the same output. Based on the output you can't distinguish what the models are. So you need one more condition so you have a unique model. You can say, given this data, what is the one model that fits? The way to get uniqueness is to get invertability. So if we can go from our Xs, we can go from our know what the noise was and calculate X by the sum, you should be able to have some way to go from X and calculate back what the noise was.

So I've just gone and written that down as there exists some coefficients I can apply to these past values of X to get out this value of W. If that's true, if both of these things exist, if I can go from W to X, I can go from X to W, then this operation we've done is invertible, has an inverse. So this is the condition to be invertible. If it's invertible then you have only one unique set of coefficients, one you chose such there they are invertible. So if you go through it, you end up in one place and you come back, you end up in the same place.

So this chapter is on ARMA models. So this is the AR model, this is the MA model. The ARMA model has both of these characteristics. So X of T is going to be based on some mean function past value, maybe some other coefficient times T minus two, and then some weight and then again this scaled version of an old weight and then some other weight and a so on. So you've got literally ARMA. So again this is our full auto regressive moving average model. Then one of the important things is that these fees here are made so that they are constrained so that keep this part stationary. These things are constrained not by stationarity because they don't affect it, but by invertibility.

This is giving you your auto regressive model and the number of coefficients, how much history you have is going to be P steps back for the auto regressive part and Q steps back for the moving average part. So you could have an ARMAPQ model, which is telling you how far in the past to look at your own output and how far in the past to look at this correlated series of noise terms.

Last section is on forecasting and it's important to understand how this works and what's missing to understand the limits of it. So you have our X of T. This is our AR1 model. We've had this kind of running example. So you should say, "Okay well XT plus one equals fee of XT plus WT plus one. So this notation here where they're putting these superscripts, this is the value of at XT plus one given this history up to point end. So that's what that superscript is telling you how much of the past history you're using that you're adding it all up.

Then the problem is we have this XN we can plug in here. But we don't have this weight. But if we replaced it with the expectation of the weight because this is a normally distributed random variable with mean zero, that's definition of white noise, the ... I'm sorry, this should be an expectation, The expectation of this is just zero. So we just zeroed this out. So we're basically taking our next time point and then running it through but we don't have any noise anymore. They also derive two step ahead forecast and then they go right to this general formula. So this general formula now it's given that history N at N plus M time point. So the N points in the future you're just going to have your coefficient multiplied by your past value. Then again these weights, you're just basically saying there's no more noise coming into the model.

So it shows this problem with forecast of these auto regressive models where you might have something happening and then you want to make a prediction and all you're doing is taking this value and scaling it. Remember that each of these is less than one. So you have a number less than one raised to some power. Each time you do this coefficient gets smaller until this whole thing hits zero. So there's kind of a limited amount of forecasting you just hit zero. Or if there's a mean, until you hit your mean function.

Okay, this lecture is going to be on ARIM models. So this is auto regressive, this is moving average and this is going to be integrated. Basically what a ARIMA model is going to be a model that when you difference it gives you an AR model. So this lecture's going to be a mix of some of the content in the chapter and also going through some of the coding examples. So for first I want to review AR models real quick just so we all start at the same place. So the auto regressive model here, the important thing is your output X depends on past values of X. So this is a thing where what happens next can be predicted from what happened previously. One example that is pretty easy to understand is the temperature is going to depend on what the temperature was.

So it's going to change kind of up and down over the course of each day. So in contrast for our moving average model, we still have the same output here at X, but now we have these new noise terms that enter at each time and we have some lagged versions of the noise. So this is a model that's depending on noise being pushed through some combination of past values. It's important thing that there's no X here for the moving average model. In the ARMA model, now we have both this noise term, this is the MA part and this is the AR part.

So you have some mean and past values of X and then some new noise coming at each time step and new values of the noise. When you say an AR model is PQ, this numbers are the order, this is how much ... For the P you think of past. How much of the past outputs you're using. For Q it's in your moving average how many of the past values of the noise you're averaging. Okay, so P is past values of X. Then Q is how many of these past values of a W you're getting. Now the other thing we talked about are models where you have a combination of a stationary part and a non-stationary part. So you have again your output XT and that's this MT plus YT. I'm going to say this part is stationary. So for the first example here, they plug in this for the U and they run through one line of algebra and you get that this delta XT, this difference in your outputs is going to give you one of these terms that survives.

You're going to get this difference of these Ys from your stationary process. So by applying differencing we've gotten down to something that is stationary. So the idea of the ARIMA model is we're going to take a model that might not be stationary, we're going to keep differencing it till we get something that we can work with. So now we can get our definition of what an ARIMA model is. Remember the ARMA, like we just said, you have these past values of X, this is how much of the output you're averaging into your model. This is how much of that internal noise is going through the moving average inside the model. We're going to add how many times we have to difference the model to make it stationary.

Now before we can talk about the forecasting, we need to review the notation for the forecasting. So if you want to make a prediction of X at the next time step based on the history, so that this is collecting these two numbers we would need. So we're going to do the history up to N and predict this step. But once you start plugging numbers and you might want to say, "Well if the history at first 100 samples I want to pick the 101st." Or if I go to the history up to 200 samples, predict the 201st sample. So you will be plugging your numbers in here. Then for this auto regressive model, we're taking our number here of XN and we're just multiplying it by this coefficient fee. You'll see here there's nothing because as we went in the last part of module four, we're not adding in that noise term.

So when we're forecasting things are going to kind of drift towards a zero because that noise input isn't pushing the process forward. So the forecasting's only going to work for a little bit. So for this ARIMA model, now we're differencing to get an auto regressive model. So we're taking our X and we're differencing at some number of times to give us Y. So now when we want make a prediction, we're going to forecast Y, this difference signal and then we can plug it in to get a prediction of X, our original output.

So if you think through what happens with this differencing, our prediction of Y at N plus M time step. So we're going to take all the history up to N and then make a point prediction M points in the future. But Y is going to be just this difference between the Xs because we difference them. So it's going to be this X at NM plus one. So this is the same time. So we can take this and then we're going to take our previous X from the time step before that we've differenced. Then you can see all these terms just get rearranged here. So if we want to make our prediction of X in a future, well we know that if we take these differences we get Y, or Y comes down and then this is the sign switch because this flip sides, but it's just algebra.

So it should be pretty a straight forward. X is still our output like it's been for the whole course, it's a thing that we care about. But we're doing this indirect thing by modeling Y and then putting the pieces back together to give us our X forecast. So it's a little more tedious bookkeeping but otherwise it's not much you haven't seen before. So if you're a solid on ARMA stuff, this is going to just be this differencing and then you can work with it and then you got to put it back together. The other thing to keep in a mind with this chapter, and this is especially why we're going to do some of these examples on the computer in a minute, is that this is a little bit different from a model fitting in kind of a data mining or a machine learning thing because it's possible to have some different ARMA models that will fit the same. So you can't tell them apart but they'll have different orders.

So it's important, it's why there's a bunch of time spent in getting to the simplest model. So the point is a lot of the stuff is just algebra and there's some superscripts and subscripts to keep a track of and it's this kind of thing. One of my best teachers used to call this stuff bookkeeping. It's not really exciting or that interesting, but all these little details have to be managed right in order to get you the thing you want, which is this prediction in the end.

Here's the first prediction they have from this auto aggressive moving average model. Let me make that full screen. Okay, you don't need to see me. There we go. Okay, so in the plot window here you can see it's making predictions for 50 time points in the future from an AR1 process. The thing to see is they have, both those red lines that are the predictions but also the idea of where the errors are and that the error is growing over time. So that's different from the predictions of what we had for just the auto regressive model in the last chapter where it's going to zero. Now it's just a line, it's just going up over time.

Now we're going to look at a very important special case of ARIMA model and this is one I've actually used a lot in my own previous work, which is just to zero couple of these out and to talk about using it as a smoothing. So let me just ... All right, so I'm going to just leave this up with the PQ and the D labeled so that when you see things come up you can look up and see what that means. So they talk about this model, this ARIMA model. They just call it 0,1,1. So you should be a familiar enough with all these things, meaning you can look at 0, 1, 1 and know exactly what's happening. That means there's no past. It's having one a noise term and then it's doing one difference. So it's differencing once, and this is the noise term, the Q. So this is going to say that your output, your XT, if you difference it you get something that's stationary. So you can difference it once and you're going to use one bit of noise from the past.

They write it in a slightly different way in the text but you can rearrange terms [inaudible 00:34:20]. But they're going to write it like this. So you get X of T minus one, plus this old value of the weight and then minus lambda, times the WT minus one. So it's a different symbol here, but this is your moving average parameter. So if we look at what the prediction is, this is going to give us a better understanding of what's going on here. So you have some number here, one minus a lambda times the old X. So it's kind of like an auto regressive model but we're multiplying it through. Then we also have the prediction of what you would have at N based on all the samples up to the previous one. So you've got these two things based on this lambda term and this lambda is less than equal to one.

So this actually is a very special case of a model that you might call an exponentially weighted moving average. This is something that's tremendously useful. So let me show you the plot of what it looks like. All right, so what you can see in the plot, it should be a little bit thicker, but there's a black line and a red line and the black line is this kind of jagged thing that's moving around fast and the red line is an average of the previous time points. So one way to think about this is as something like some kind of average of the previous points like a filter, but it's different because if you compare it to a moving average model, the moving average model has this order of Q. So the moving average model has a fixed amount that it looks back in the past and then it doesn't forget about it. This model looks at the entire history of the whole past. So the exponentially weighted moving average, it's like an average but it's an average of everything that you've ever seen.

So you can get a similar performance of a smooth trend, but it's not predicted the same way. It depends on everything that ever happened. This is usually good. But one of the things that that means is that if there's some shock or some weird thing to your system, it's going to last forever. Another way we are thinking about it with this I, and I mean it's integrated. So you're looking at the whole history, you're integrating everywhere it was to figure out where it's going to go. That's different than just averaging the last 10 samples.

Okay. So this is the example 5.6 in the book. This is data from the US census. You can see it's going up and if we look at this auto correlation function, you can see it's highly correlated with what it did previously because it's always going up. So this is a very strong trend. We want to get rid of that trend so we can get a better idea of what's happening from quarter to quarter. So the way we would do that, I'm just going to paste in the code from the website, but it's basically just doing the diff and then running all the same things. So now you get something that's a lot more interesting, you get these differences and you can see that compared to how it was on the previous step in the previous quarter, in the previous three months in this example, you can see sometimes it's coming up, sometimes it's not making much, much of a change. So this now, we've gone from our X being this gross domestic product to being our Y.

So we did a difference of one to the X. So now we're seeing the differences in what it's doing now versus where it just was. Then at this point what they basically do is you're getting back from the code here, this auto correlation and the partial auto correlation and they just decide it looks like that it's a moving average over two. So then they want to fit this model that's auto regressive integrated, ARIMA, right? So 0, 1, 2. So there's no auto regressive part, there's a moving average of order two and you difference it once. Then you can just call the code that fits it and say, "I want to fit a model with these parameters," and it'll fit it and plot and you get this diagnostic plot of the residuals that lets you see if it's a good model or not with a QQ plot and so on. But the point of this whole chapter and stuff is that you really need to know what these kind of things mean so you know what numbers to plug in here.

There's also, this part on fitting models talks a lot about these things that are more, it's not a hard set of rules algorithm, it's kind of like a troubleshooting guide like, "Do this, look at it, do this, look at it." Then that it's kind of more of a heuristic. It's not like saying you must do it this way, but it's like here's some general ideas. Make these numbers be a small, then gradually increase them. Otherwise you could fit a model that has too much history if you do too much differencing, you're going to introduced dependencies, which you might remember from when it was discussed in the previous modules. So don't just difference it all the way up, but just build it up in stages is kind of the point of this whole thing.

Having done that, it then goes on to say that this partial auto correlation here might actually be of order one. So then you can change that and fit it. Now you have two different models that fit almost the same but are telling you a slightly different story about what the model means. I don't think I've covered the QQ plot yet in this class, but it's definitely worth looking it up as kind of a review from regression analysis. But basically you want it to look like it does here so that you are getting things that are actually distributed normally, which is what you assumed and what you would expect.