

# MITx 15.455x Mathematical Methods for Quantitative Finance

## Recitation 3

September 23, 2021

### Module 3: Forecasting

Let's talk about forecasting. We saw in lecture that forecasting is an application of conditional probability to time series models. So we think of the time series as generating the data. And we ask at a given point in time, given all the observations that exist up through that point in time, what can we say about the future? We might want to know what happens one time step ahead, two time steps ahead, or in limit as the number of times that goes to infinity, but the key thing is the break between past and future. So we take the present time to be just after we've made our final observation. And we're first forecast period is one step ahead. And, as I said, the key tool is in addition to our previous ones with expectations, linearity, algebraic substitution, recursion. We had a new one, which is conditional probability. In this setting, conditional probability simply means that for any variable that was a random variable in the past that has been realized, it's no longer a random variable. Now it's a scalar. Now it's a number. So those things that show up in the recursion that we previously thought of as random variables, in this setting, when we take conditional expectations of our defining equations or equations of evolution, then they take a different character, because they're already known. They're no longer random variables. So we looked at an example in lecture. Let's just do a couple of examples together. It's the same structure. They're just one or two things it might be worth paying attention to as you look at the blind forecast of your own. Some of the things that might have look special for the case of the AR1 are quite general. So let's take a look at a model.

### ARMA(1,1)

Let's look at, say, an ARMA(1,1) which is this form:

$$x_t = c_0 + c_1 x_{t-1} + \sigma z_t + \phi_1 z_{t-1}$$

So I'm going to keep the notation  $\sigma z_t$  for our shock for innovation in the period. You could give it another name for its coefficient. Remember that  $\sigma$ , in this case, is not the standard deviation of  $x$ . It's just a scaling parameter for the random shock  $z$ .

So the first thing we'd like to do is compute the mean value for  $x$  and simplify our expression. So we use our usual trick.  $\mu$  is the name will give the expectation of  $x_t$ . And then we take expectations on the right hand side and apply linearity. So we have this is going to be:

$$\mu = E[x_t] = c_0 + c_1 E[x_{t-1}] + 0 + 0$$

$$\mu = \frac{c_0}{1 - c_1}$$

And you should recognize that expression from our AR 1 model. It's exactly the same expression. And therefore, we can substitute and rewrite our equation as a bunch of things that are grouped together in such a way that they all have 0 mean. That is to say, if I substitute for  $c_0$  and say it's equal to  $\mu(1 - c_1)$ , then I can write this as:

$$x_t = \mu(1 - c_1) + c_1 x_{t-1} + \sigma z_t + \phi_1 z_{t-1}$$

$$x_t - \mu = c_1(x_{t-1} - \mu) + \sigma z_t + \phi_1 z_{t-1}$$

Now the next thing we want to do is put things in forecasting form. So what I'm going to do is I'm going to shift  $t$  to  $t + 1$ . Everywhere where I see a  $t$ , I'm going to write  $t + 1$ . And that puts the future values on the left and the present and past values or the known values on the right hand side. So for one-time step ahead, this makes exactly the split that we want to have:

$$x_{t+1} = \mu + c_1(x_t - \mu) + \sigma z_{t+1} + \phi_1 z_t$$

So what's our forecast? So our definition for our forecast is going to be the conditional probability. So our first forecast is going to be:

$$f_{t,1} = E[x_{t+1}|x_t, z_t, \dots]$$

That is the forecast made at time  $t$  for one-time step ahead in the future. And there are different notations for this. So what you really should rely on is not the notation, but reduce everything to expectations. So this is the expectation of  $t + 1$  given everything up through that point, which includes  $x_t$ ,  $z_t$ , and any previous failures, but those don't show up in this equation on the right-hand side. So let's compute the expectation. And what we noticed is that the only random variable on the right hand side is  $z_{t+1}$ . That's the only thing that's random. So  $x_t$  is known. This  $z_t$  is known. There's only one random variable, i.e.  $z_{t+1}$ . And we take expectations, the expectation of  $z_{t+1}$  is going to vanish. So what do we get? We're going to get that the forecast is equal to:

$$f_{t,1} = \mu + c_1(x_t - \mu) + \phi_1 z_t$$

So we just do the calculation. It's exactly the same as the expression above, except for this  $\sigma z_{t+1}$ , this random term that dropped out, because it has zero expectation. So let's take a look at the forecast error. So the first forecast error for one-time step ahead is going to be:

$$e_{t+1} = x_{t+1} - f_{t,1}$$

That is the forecast error as defined as the difference between what we predicted and what actually happened. What we'd like to do and the way in which we have this definition for the forecast being optimal is we're going to minimize the mean squared forecast error. That is going to minimize  $e^2$  in expectation of all possible things that would happen. Now we could use other kinds of loss functions. Those will depend on the settings, depending on the economic values. So it's just a question of doing and minimization, but in this case, we have this basic result that this will be our expectation, but let's compute the forecast error. So the idea is that what we really want, in general, if you want to drive this and check, what we really want is we want a predictor for  $f$ , that some function of all the previous observations, in this case, just  $x_t$  and  $z_t$ , we'd like to be a linear function. And when we find the linear function that minimizes the mean squared error, we find that it's this conditional expectation. So let's compute this quantity. And we'll see that there's a really nice structure that shows up.  $x_{t+1} - f_{t,1}$ , you notice that most of the terms are common.

And therefore, the mean squared forecast error is:

$$E[e_{t+1}^2] = E[x_{t+1} - f_{t,1}] = E[(\sigma z_{t+1})^2] = \sigma^2$$

Now what about looking at multiple forecasts? So we'd like to go for future horizons. One-step was easy. So the general rule is that when we want to do it two-step ahead forecast, we're going to shift everything. We're going to add one more. So we write an expression for  $x_{t+2}$ .

So here's the general procedure. Keep doing the recursion. Keep substituting in back until all of the  $X$ 's on the right hand side of the equation have a time index of  $t$  or earlier. The  $z$ 's can be later. That's OK, because we're going to take their expectations, and they're going to vanish. So we might have some possibly unknown  $z$ 's. But the basic idea is once we go beyond one-time step, how do you know what to do? It's easy. You're doing  $x_t + h$  for horizon  $h$  on the left hand side. And on the right hand side, you do probably  $h$  or  $h - 1$  recursive substitutions of the defining equation until we can express  $x_t$  plus  $h$  in terms of  $x$ ,  $x_{t-1}$  minus 1, and so on. Those are all known quantities. Then we take conditional expectations. We have our forecast. We compute the forecast errors. We take expectations of their squares. That's easy as well, because it's just

the same quantities. Now we have the known and the unknown ones clearly delineated. And we get our results for our mean squared forecast error. And there are two really important properties to keep in mind for our solution. One of them is the forecast error has zero expectation. And the forecast error is orthogonal to the other variables, to the other to the other predictor variables.

## ARMA(2,2)

So let's take a look for another example Let's take a look at an ARMA(2,2), shall we? So now when you're given numbers for the forecasts or particular horizons, you can either plug them in right away or you can leave the parameters general. That generally makes it easier to check your math and find any sign errors that might be there, but either way, whether you substitute indefinite numbers for the parameters before or after shouldn't make any difference. When you're doing forecast, you do want to pay attention though to the initial conditions if you have to bootstrap your process, because if you're asked for a forecast one or three or seven steps ahead, you'll need enough data to get the process started before you can generate the recursion. The recursive techniques that we have that I'm writing down here for the forecast are technically appropriate for the case, where we go infinitely far into the past. Even if the series did exist an infinitely long time, we wouldn't have an infinite amount of data. So we do need to make sure we have the initial conditions set to get numerical answers, but the basic rules are compute conditional expectations after writing our defining equations with variable to be forecast on the left hand side and only known observations on the right hand side.

So let's do this for ARMA(2,2). it's little bit more complicated at least initially. So it looks like:

$$x_t = c_0 + c_1x_{t-1} + c_2x_{t-2} + \sigma z_t + \phi_1z_{t-1} + \phi_2z_{t-2}$$

And you can see how it would go for a general ARMA(p,q). I'm going to write it in concrete form so we don't have summations running around. And when we shift, it gets a little messy keeping track of the indices, but you can do it. And you can take a look at the literature as well for that. So what's our expectation in this case?

$$\begin{aligned}\mu &= E[x_t] = c_0 + c_1\mu + c_2x_{t-2} + \mu \\ \mu &= \frac{c_0}{1 - c_1 - c_2}\end{aligned}$$

And you can guess how this generalizes for the general ARMA(p,q) model, we get the generalization of this expression. And then we can write our equation as:

$$x_t - \mu = c_1(x_{t-1} - \mu) + c_2(x_{t-2} - \mu) + \sigma z_t + \phi_1z_{t-1} + \phi_2z_{t-2}$$

Let's do some forecasting, shall we?

You should notice that each of the terms that I've written down is a way that it has 0 mean. That just makes it a bit easier to see what's going on, to see what the dynamics are, and to do some calculations. You can also, if you want to clean it up, you can find a new variable  $y$  to be  $x_t - \mu$  and shift things back. So you can do any rescaling things you like. The end results are going to be the same if they're just redefinitions of the parameters and rescalings of the variables.

So let's take a look for time  $t + 1$ :

$$x_{t+1} - \mu = c_1(x_t - \mu) + c_2(x_{t-1} - \mu) + \sigma z_{t+1} + \phi_1z_t + \phi_2z_{t-1}$$

So that one is pretty easy. It looks like the expression that we just did before for the ARMA(1,1). That is everything on the right hand side has an index of all of the  $x$ 's are taken at time  $t$  or  $t - 1$ . And the  $z$ 's are mostly in the past. We have these two with coefficients of  $\phi$ 's both in the past. And this is the only one that's still a random variable, but at time 2, things change. So let's look one more time step ahead. Let's contrast a little bit. So if we write this as:

$$x_{t+2} - \mu = c_1(x_{t+1} - \mu) + c_2(x_t - \mu) + \sigma z_{t+2} + \phi_1z_{t+1} + \phi_2z_t$$

So all I've done is I shifted  $t$  to  $t + 1$  again.

So I have two of the terms on the right hand side involve the present or the past. And that's OK. Those are known quantities, but I have these other expressions. So I can't just directly compute the expectation on the left hand side. Now I could take the previous work I did before and substitute that would be fine. But if I want to see the general approach for doing it at a particular time step, what I'm going to say is that the expressions  $\sigma z_{t+2}$  and  $\phi_1 z_{t+1}$ , I can leave alone, because those are both fine. These just involves  $z$ 's. And I know their expectations, but I don't want the recursive structure for  $x_t + 1$ , because remember this is saying that the value two days from now depends on the value tomorrow, which I haven't yet observed. So the value today, the value yesterday, those are all known, but rather than expressing it this way, I'd like to do my recursions, get it out. And then I can take clean ordinary simple expectations and get an answer. So let's do that. And we just need to substitute one more time and put that in. So which term are we going to substitute— we're going to substitute in? We'll have :

$$x_{t+2} = \mu + c_1 \left[ c_1(x_t - \mu) + c_2(x_{t-1} - \mu) + \sigma z_{t+1} + \phi_1 z_t + \phi_2 z_{t-1} \right] + c_2(x_t - \mu) + \sigma z_{t+2} + \phi_1 z_{t+1} + \phi_2 z_t$$

So what have we got? So I now have an expression for  $x_{t+2}$ , where on the right hand side, everything depends on known quantities. So we put it in. We turn the crank. We compute the expectation two-time steps ahead. That will give us our forecast today for two-time steps ahead. When tomorrow is realized, that forecast for what will then be one more day ahead is going to change. And how will it change? It will be updated by the new observation. So if today is Monday and I'd like to know about Wednesday's weather, I would make a forecast today. Tomorrow I'll have a new forecast for Wednesday. And that will change. It will change by my knowledge of what happened on Tuesday. But from a mathematical point of view, we separated things into the form that we want. We can first take the expectation with the appropriate conditions. And second, we can compute forecast error in advance. And we can compute the mean squared forecast two steps ahead. And so that's looking ahead for what our expected forecast error. The reason that that's important is that we want to think not only about distributions. What's the exact value? Because that exact value probably won't be realized. It gives us characteristics of the full distribution of the actual outcome. And it lets us have a sense as to how we should evaluate our forecasts. So that's an expectation. Once Wednesday is realized, then we'd like to go back. And then our  $X_t$  plus 2 will be a known quantity. We can compare it with our forecast. And, of course, this will differ by some amount, but if we'd like to improve our forecasting techniques, our forecasting methodology, and our forecasting quality, what we do is we look over time. We take a large collection of forecasts and outcomes. And we study the statistics of the forecast errors. So we use forecast errors both in expectation going forward and to assess forecast quality and look for improvements going backward once we've collected the relevant data.