MODULE 1 UNIT 4
Video 1 Transcript

IN COLLABORATION WITH getsmarter™

Module 1 Unit 4 Video 1 Transcript

NIR VULKAN: Hello everyone. I have developed principles that can be used to assess algorithmic trading models. These principles are presented in two sets of four. The first set of principles relate to the guidelines for building an algo model, and the second set relates to evaluating an algo model. These principles will be revisited in the course, specifically in Module 4 and 5, and we will see the rest of the material that we cover in 2 and 3 will be used for that.

Let me first take you through the four rules for building an algo model. The first principle is to understand how is your model making money. Money can be made in three ways: You can make better forecasting, better prediction, of what's going to happen in the market, or you can gain an advantage by having a better execution strategy, or you could build a model that takes advantage of your ability to manage risk.

Most of the models we see in here fall under A. Most of the funds that we will see in this course, in this programme, would be trend-following, taking advantage of strategies like that. This has to do with trying to predict what will happen tomorrow. You basically build a model that looks at the data that you have, and you try to make a prediction of what will happen tomorrow. And what we will see is, what you do is, you make lots of prediction in lots of different markets and hopefully most of them go your way. This is what we mean by forecasting. However, there other models that make money through execution; most of the very high-frequency stuff that we see come from that, and we have, we will see, we will have a video tutorial by Stefan Zohren, that talks specifically about limit order books and how you can make money that way.

Finally, you can make money by having an advantage in the way you manage risk. This is really to do with large funds. I suspect most of you guys who are maybe starting in this area wouldn't really be able to do that; this comes really with size. Nevertheless, there are strategies, you can build models that do that, and before you start building any model, you have to understand which of these three you belong to.

The second principle I name “Identify and quantify the opportunity, and can you explain it?” This is really important because, for example, if you're talking about trend, can you explain it? You can explain it in terms of the psychology of the masses, if you like, yes? We've seen this in behavioural finance; the things that you can explore using that, you should be able to explain, because if you can't explain it, not only you will have difficulties raising funding or selling it to investors, you also take the risk of that you’re too close to the data, and what you're doing is data-fitting, and data-fitting is sort of the enemy if you're trying to build system, because you might get things that are, that look really good on simulation, but may not work. And so, try and explain it, try to be as clear as possible, being able to explain it to non-technical people. What is the opportunity? What kind of abnormality in the market you're trying to capture and take advantage of.

The third principle is “Test and verify”, and also use a meaningful out-of-sample, be disciplined, and be honest with yourself. You will see in Module 3 and 4, where we go through the process of building a model and fitting a model, we will explain in details what we mean by out-of-sample or statistical verification. And, you will hear from many of the experts here – people who’ve been in industry for many years – how important it is that you take your model and, whatever, how much you love your model, how great it looks in simulations, the most important thing is how the model performs on what we call the out-of-sample, i.e. on a number of years or months of data that are unseen, unseen by the model in the simulations before. This is the closest we can get to testing what the model would look like in the future, yes? Because this is data that the model has not seen before.

Now, we will go through this in detail in Module 3, but there is a lot of discipline here, because what happens often is you find a model, you kind of fall in love with the model, it performs really, really well, you can explain it and everything, and then when you run it out-of-sample, it doesn't work. And, what you want to do is to say, “Let me go back and change the model a little bit, maybe it will work”, and then you do it again, and again, and again. This is what we call burning the out-of-sample. If you do that too much, you will probably find something that works, but it would be coincidental. This is where discipline comes in, and discipline being being honest with yourself first of all – before we talking about being honest with investors, obviously – being honest with yourself, being able to say, “This did not work, I would have liked it to work, but it did not really work in the sense that it, when given it an unseen data, it didn't work”.

Finally, our fourth and last principle is to connect and to build the model, connect and test the model real-life data, and to watch out the performance, and particularly watch out what we call slippage and unforeseen costs.

I have seen, over the years, many nice models that look really good in theory, that maybe have held out-of-sample, but once they get connected, once they start trading, you realise that the people who have developed them have not really thought about the trading assumptions, and particularly the slippage involving getting these models executed.

We will see that lots of technical trading models are based on what we call “limit orders”. Limit orders is: You express to the exchange a price at which you willing to buy or to sell the underlying assets that you’re trading. Having expressed that, doesn't mean you immediately get filled at that price, and the process is called slippage, and this can become very, very significant.

We will see a number of techniques of how you can deal with that, and how you can try and estimate these things as realistically as possible. Now, this is particularly true for those of you who come from maths, and physics, and economics, and so on, but have not really worked in markets before. This is also one of the reasons that the funds that tend to do better, the algo funds that tend to do better, have a nice mixture of people from the markets, and people who are finance practitioners, and academics or theoreticians who work on the models, and this is because this is a really, really important part to get right.

With all the computing power we have now, if you get it wrong, if you’re kind of being a little bit optimistic about how easy it is to execute the model, then your computers, your simulations, will show you really, really good performance, and when you start trading it for real, the model still does well, but there's a little bit being eaten away from it through the actual cost that you haven't really modelled properly, and then you will see that it takes a long time to realise the model is not so good.

And so, it's really important to connect, to be patient, to really get – as much as you can – the assumptions right, and then once you've connected the model, once you started either paper trading or really trading the model, or trading with real money, continuously monitor the performance of the model and compare that with what you have in your simulation, and see that the two match.

I'm now going to introduce the four rules or principles for evaluating an algo model. This is hopefully useful for you guys who have not interacted, who work in this business, but have not interacted with algo trading firms before. So, the first, I think it's self-explanatory: If it's too good to be true, it's probably not true. You will have people coming to you with flashy degrees, with a sort of army of PhDs in physics or other discipline, or you know, rocket science, and you know, promising doubling your money every year, year after year.

That's probably not true. It is possible, yes, to double your money for a couple of years in a row by taking really, really high bets, but these are the strategies that will then, when they don't work, will also lose you all the money that you have. So, I mean, this is just basic common sense.

The second principle is, can they – can the people who are coming to you, can the algo fund people – can they explain to you, can they explain to you what the model does using basic economics? This is very important because if they start to say, “Well, you know, it's all AI, machine learning, neural network. You wouldn't get it”. I would be careful with these things. It is true that there are very complicated techniques in machine learning and in AI, and it's also true that as an investor you don't need to understand exactly the formula that people use.

But these models, as we've seen, as we will see in this course, they really capture basic behavioural finance patterns, and so whatever the model captures has to do with that thing, so it's really important that they know to explain it. If they know to explain it then they sort of know what they're doing, and that is what you are trying to estimate here. And this is particularly true with this, as I said, with quantitative model, if you are not a quantitative person, don't be afraid. They need to be able to explain at least the principles of what they're trying to capture, yes, in terms of basic economics.

Now, the third principle is perhaps a little bit more technical, and this is, did they overfit the data, okay? Very complicated models can do really, really well on simulations, but they tend not to hold out-of-sample. Now, this is not something that you as an investor can have direct insight into – they wouldn't tell you obviously, they'll tell you they did it – so you should try and engage with them as much as possible.

Having done this programme, you would have played yourself, you got to build a little model, you will see what an in- and out-of-sample is, so you get some insight into it, and you will be able to ask these guys how they did that. What percentage of the data was unseen? Did they look at good and bad years when they check their models? And so on. Make sure they give you simulations that include drawdowns as well. So, if they only show you, for example, the last couple of years, and these last two years look glorious, you say, “That's wonderful. Can I see back simulations, you know going, you know ten years?” And so on. And particularly what happened in 2008? How would the model have behaved in a particularly challenging a period of time?”

So, asking these kind of questions and challenging them a little bit, can give you an insight into the degree of overfitting of their model.

So, four really reiterates what I said in the third principle, and this is, was the model tested in different market condition, and how would have performed in bad times? In bad times you know, you could, if this is a CTA, for example, you will have access – it's very easy to find online – how other CTAs have performed in certain years, and you can look, let's say 2015 was particularly bad year for CTAs, ask them how their algorithm would have performed in 2015, or has performed if it was live already, if it's not simulation.

So, challenge by looking specifically at difficult times, and remember difficult times relative to the assets that they are trading, okay? And relative to the category that they are in, because if it's a trend model, for example, trend models tend to do better at times when volatility-based models don't do so well, and volatility-based models sometimes do really well in years that a trend does badly, so don't be impressed if they have a volatility trend model, a volatility-based model, that did really well in times when markets were very volatile. So, try and understand, you know, what they do. Try and then find the relative category, and then look at how the performance have been over the years and, in particularly in bad times, because if the model takes more risk – cause one thing it's very difficult for you to see is what level of risk the data, you can ask, but it's very hard to see – and if the model takes a little bit more risk, it's more leverage than, let's say, the industry in general. It will outperform the industry in the good years, but it also would underperform in a bad year, so make sure you look at that and have, through that, some understanding of the risk that they are taking.

These are the two sets of principles which I hope you'll find useful and you'll be able to remember by the time you finish this programme.

If you would like to review any of these sections, please click on the relevant button.