



MODULE 4 UNIT 1

Video set Video 1 Part 2 Transcript

Module 4 Unit 1 Video set Video 1 Part 2

Transcript

Do you think there is some mileage in markets with interesting data sets, for example, social media?

EWAN KIRK: There is a hope, a desire, maybe a fantasy about big data and new data sets, and some of these are fascinating and interesting. Shipping data, credit card data, the old favourite satellite photographs. I like lots of new data, and I like big data. But, we should always remember that finance has always been a big data problem.

It was a big data problem when they had too many ticker tapes in the stock exchange in the 1920s and didn't know how to store them. It's been a big data problem. And we've always had big data. There is an enormous amount of data in finance. This new data, I'm sort of naturally sceptical. And I think it's a good way to be as a scientist is to be naturally sceptical. If it's too good to be true, it probably is, and if this data set is going to revolutionise finance, it probably isn't.

One I liked use is satellite pictures. Isn't that really interesting? Well, yes it is. But, I kind of like to be able to trade when it's cloudy. And when you mention that to satellite data providers they go, "Oh, yeah. Well, maybe you couldn't trade on that day," or something. Also a lot of this big data, which it depends whether or not the big data is true data that just couldn't have been caught any other way. Shipping data is a really nice piece of data. There's an enormous amount of it, it's lasted for a long period of time, it tells you something interesting that you really couldn't pick up somewhere else.

But I'm not sure about credit card data. Does it help? It helps if you trade US retailers... probably. But, does it really help? Now, you know, let's think about, I've got the credit card data and I find out that more people are buying food in Walmart. That might be something I would find out from that data. What do I do with that? I mean, yes, of course, it's interesting, but maybe they're buying more food because Walmart's got a sale. So therefore Walmart's profits are going down, not up. It's not as simple. The transfer mechanism between this data and a tradable strategy is challenging. I'm not saying it's not true, and I'm not saying that there's not a lot of interesting data out there and it will have some effect on the way that we trade markets, but I don't think any of it's revolutionary.

In your firm, do you start with the data or the model?

EWAN KIRK: I think we begin with the data. Coming up with a model is not an easy process to regiment. But in general, everything starts with the data and you have to make sure that you have the data. So what does that mean for us? Well, that means that we have to be able to have real-time or near real-time data for everything that we trade. Why do we need that? So that we can make contemporaneous "snaps", as we call them, through markets at a specific time so that the data all lines up. Now, if you don't have real-time data, you kind of, you don't know where everything is and you can deal with that, but that just means that you have to delay your signal or lag your signal by a day because you don't know exactly when you got the data.

So clean data, accessible in an easy way with a rigorous and regular API is absolutely critical. Then, there's a piece of magic and that's a hard thing to do. You just know when

one of our researchers is just in that zone, and she just comes up with a new model every week because you're just in the zone of doing that. And, I can't explain that. It's like, it is a little bit like writing a song, it's a creative act, and it's hard to explain where that piece of creativity comes from. Often, it might be a mathematical technique which we have discovered, or come up with, or learned about, which is now applied to the data. Or, it may just be a moment of serendipity where you say, "This is an interesting effect". Or, and this often happens, "This model does this, but actually, what it's really doing, is this, which is different," and adding that model into it.

So, I wish there was a process. If there was a process, then it'd be easy. We could just start people on the conveyor belt and they would just get to the end. That doesn't happen. There's also... you have to be very, very careful, because if you want your significance level to be less than 0.05, well, that's okay just try 20 models, one of them will work. And the more models you try in the data, the more you're going to get a model which might actually work just by chance.

That process of coming up with the model is hard and difficult to explain. Often it's a technique. We've got a new technique, maybe that'll be something... a different filtering technique to filter time series, maybe it'll be a different forecasting technique, maybe it'll be something cool like recurrent neural networks. There's lots of things that you can use. At that point, you get to something which in-sample, and remember, of course, we're all in-sample at this point, you can maybe say there might be something here. And then the really hard work starts.

This is the end of Part 2. Move on to the next video for Part 3.