1 - Interview with Tammer Kamel Part 1  
  
I'm Tucker Balch, we're here with Tammer Kamel, founder of Quandl. We're going to dive now into details about how the Quandl platform works. So let's start with sort of a big picture, imagine I'm a hedge fund and I'm interested in working with Quandl. How does Quandl fit in to my work flow? >> Right. So in any hedge fund where quantitative thinking is some or part of the investment strategy, then at the very start of that process, is good, high-quality data. And that's where Quandl starts and finishes, in a sense. We view it as our mission to get the data that professional investors need into their hands and into the format that they need it in. >> So Quandl offers a hedge fund literally dozens of mechanisms to pull that data. Whether it's a low level API and they can pull the data right into a database that they're maintaining themselves, or into Python or R, or MatLab or Stata or any analysis tool they might be using. But we consider our job, mission accomplished for Quandl, if we quickly and painlessly get the hedge fund the data they need into the system they want. >> Now presumably, this includes a combination of historical data and live data. Are all of your feeds historical as well as live? Or are there some that are special in some way? >> Right, so all of our data is historical for sure, and updated daily. Live data is a little bit in the future for us, it's not available on the platform as of this moment. >> So I could imagine a flow something like this. Lots of people are on a daily cycle. I think that's a majority of what people do. You might read historical data to inform your model or build some sort of model. But after market close, you wait some period until the market data is available and then whatever data one might pull from Quandl, do some number crunching overnight and put in orders for what you might do in the next day. Now what time does your data available and what mechanisms do people use to pull it? >> So what you described is really a canonical use case for Quandl, right. A fund or an analyst will take the historical data, build and calibrate a model, and when they're comfortable that it's ready for action, then they're going to start executing on it. So Quandl market data is updated pretty quickly after the close, within minutes or depending on the exact database, we're talking about ours so. But definitely in time to run today's numbers through your model to get the output and know what actions need to be taken on the next trading day. >> So I love to learn details about how people build complex financial systems. So where do your machines live? >> So Quandl is quite modern in that respect. Everything lives in the Cloud. We're customers of Amazon, like so many others are. At the heart of Quandl is a modern, no SQL database. And that's where the 15, 16 million times series, few billion data points live there. On top of that, we've built an API. And then the Quandl website itself, interestingly, is itself just another client of the Quandl API. If you look at our website, you will discover it's perfectly orthodox. It's a standard Ruby on Rails website, built very conventionally, it's just another client of the Quandl API. >> Why the choice of no SQL? >> So because of scalability, right? We really need to be able to deliver lots of numerical data to lots of people very, very quickly. The number of data points in the Quandl database is in the order of a few billion now, but we expect that to be 10 or 20 or 50 billion within a year or two, and at that point, conventional relational databases get kind of difficult to work with. >> So the glue between the various components is Ruby on Rails? >> It is, indeed, yeah. >> So it's no SQL and Ruby on Rails, and that's it? >> That is it. Quandl is, by software engineering standards, pretty plain, vanilla, simple implementation. Where Quandl, where we sweat day-to-day is usually in maintaining the accuracy of 12 million data sets. The technical infrastructure is, not to say it's trivial, but it's conventional best practices software engineering. >> Getting back to the data, tell me a little bit about timestamping. Let me tell you what I'm getting at. So let's just consider market close data. So the price for a particular stock essentially is stamped at 4 o'clock, so one potential timestamp for that data is New York time, [LAUGH] one potential timestamp for that data is 4PM, but maybe I can't observe it until 4:05. That's just one example. But I'm curious, does that sort of time stamping feed into what you all do? >> Yeah, that's actually a really good point. I suppose what you're getting at is the risk of creating a look-ahead bias in your analysis, right? >> Yeah. You know what, I can't claim that Quandl has any sort of panaceas for that sort of thing. We time stamped the data, of course. And the meticulous, diligent analyst should be aware of this kind of thing. And they'll see that this database is the 4PM closing prices, but- >> This more likely did occur with things like earnings reports. >> Yeah. With those sort of things, we reach to the vendors and the publishers for this, but they are acutely aware of that, so you'll see. Take for example our earnings data. They're both as reported to the point and time data and revised data. So those kind of blatant risks of the look-ahead bias, those are well mitigated. >> When you look at your various vendors who are providing you data, what sorts of flaws do you see in that data and what sorts of steps to you have to take to mitigate those flaws? >> It's interesting, there's no such thing as a perfect data set in this world, right, and we're of course strongly motivated to get as close to that ideal as possible. The way we've done it is kind of simple, but also a little bit original. And it falls from the mechanism we've set up, where consumers of data have a direct relationship with the publishers of those data. And indeed, the publisher of the data's livelihood depends on keeping that consumer satisfied or happy. And so unlike sort of many other data platforms, where there's a lot of opacity between the end user and a lot of layers too, between the end user and the actual publisher. In fact, you might in many cases, not even be able to reach the publisher. Quandl has this direct relation. So what you find is that if and when there's errors in the data, our publishers are very keen to correct it, and fast. Because they've got 5 or 10 or 50 customers who know it's them who are responsible for it, and they're keen to keep these people paying for the service. Right? So this is an amazing, this is a reality that happens in many marketplaces. When you start removing the distance between the end consumer and the actual producer of the product, you get this quality improvement. The other thing that, of course, affects that is good old healthy competition on a marketplace. We actually, right now, for some databases, we have two or more vendors offering competing products and nothing drives quality control like competition. >> Okay, great interview. Thanks very much. I'm sure the students out there will find this fascinating. Thanks for your time, I really appreciate that. That's the end of this discussion with Tammer Kamel, and we'll see you online.

2 - Interview with Tammer Kamel Part 2  
  
Hi there, I'm Tucker Balch. I'm here with Tammer Kamel, he's founder of Quandl. We're here at Quantcon, which is a conference organized by Quantopian. So I want to talk with you now about your experience advising hedge funds. I guess you have some experience actually managing money. And students in the class are really interested to hear about real live strategies that succeeded or failed or lessons learned. So what is jump diffusion? >> The jump diffusion was a model to try and capture the sixth sigma event. And the idea here was, it was a smart idea, and I can't take credit for it entirely. But the idea here was if you mix normal distributions and drew from them at different probabilities. So for example, you had a distribution with volatility of sigma. >> I think I know what you're talking about, but I just want to review it, in case folks there, so they can follow. So you might have a distribution of returns for one strategy or one stock. >> Right. >> And so you're talking about the distributions for maybe different strategies or different stocks and then potentially combining them in some way. >> Yeah, exactly. >> And everybody assumes those distributions are Gaussian, right? So the bell curve- >> Right. And so, Gaussian distributions are so delightful to work with, right? >> [LAUGH] >> So, the problem is that reality doesn't intersect with them that much, right? But one nice thing you can do is you can actually simulate reality a little bit better if you use not one, but two Gaussian distributions. And as your random variable moves, you draw usually from the first normal distribution. But then with some random probability, you draw from a distribution with a much higher standard deviation. Thus simulating these rare events and creates the fatter tails, and is great for risk management. It's predicted power is minimal, because it's random so you don't. But for risk management it really gives you a much better sense of the fact that- [CROSSTALK] >> So you might have a current portfolio and do some sort of Monte Carlo simulation to see how risky that portfolio is. >> That's exactly the use case. >> Gotcha, gotcha. Okay, okay. So in one of your earlier careers [LAUGH], you advised hedge funds. What sorts of strategies did your company generate and advise the hedge funds on? >> So I graduated from an engineering program at the university in the mid 90s. And I always call this sort of the golden age of quantitative financing. It was an era when Wall Street was plucking young vulnerable engineers like myself, >> [LAUGH] >> And having us mathematically model markets. And back then there was all kinds of really fat opportunity. The first thing we used to do was- >> Anybody with a computer could make money, right? [LAUGH] >> Well if you could do a bit of math you really could, right. There was this thing, we made a lot of money on was yield per arbitrage. >> If you did just a [CROSSTALK] model the swap curve, for example, with a two or three factor model. And you would discover, a, that at any given time, some deviation between a sensible looking yield curve and what was- >> So a particular asset would be currently priced in a such a way to be contrary to what the yield curve would predict? >> What it would come down to is it would be mis-priced relative to its peers on that yield curve, right? And backtesting would tell you, would confirm that time and again, that anomaly would revert back to something more normal, right? So you could make a lot of money by taking these reasonably sophisticated yield curve positions. Where you were long at two points of the curve, and short at two other points of the curve, and you get these funky butterfly kind of trades, right? But they'd be sort of neutral to the level of the curve, and even neutral to the slope of the curve, but there's these third and fourth factors that would be in revert all the time. >> So, to take a short position in a fixed income, you would have to be a credit swap of some sort? Or how would you do that? >> Many ways you can do that. Credit swap certainly does it. There was no such thing as a credit swap when we were doing this. You could short a government bond, borrow it and sell it in the repo market? >> I just think of with the way you can trade today electronically, you never really think about the borrowing and the reselling. >> Right we used to worry about that stuff. Yeah repo rates, and all of that. But yeah, that's been extracted away to some extent, right? But yeah, that's how we used to do it. >> But, so you talked about looking at yield curves. Did you all do anything in equities? Or was it just primarily the fixed income? >> We did, also in the 90s, stat arb and even simple pair trades were very effective trading strategies, right? There was a lot more inter-stock volatility. A lot less correlation between the stocks, at least until the late 90s and the crash. And you could even do more sophisticated stuff like the stat arp stuff or eigenvalue PCA analysis on stock markets and do sort of payer trading strategies on these. >> So a lot of that stuff doesn't work anymore. >> No it doesn't. >> [LAUGH] >> That's why I started a data company. >> I see, okay. Anyways, I wanted to ask to what do you attribute the evaporation of alpha and these kinds of approaches? >> I think there's a lot of interesting things going on, right? There's no question that the number of participants executing similar strategies is a factor. Volatility is low these days, right? And these strategies thrive in high volatility environments, right? Like the 90s was a much higher volatility environment, even in a bull market. Another big advantage we had in the 90s was the phenomenon of the day trader, which was for lack of a better term, incompetent trading. >> Right. >> Right, which would create all kinds of wacky, if the day traders woke up on a Tuesday morning and they were all in love with Ebay for example. Then everybody just went and bought that, and it would create a mispricing in Ebay, right? But what you don't have irrational actors in the marketplace, it makes these relative value pair trades are harder to adapt. >> And I guess at that time maybe the day traders were a higher percentage of the volume, right? >> It was definitely a large fraction of the volume, or a meaningful faction of the volume. >> Taking food from children's mouths. [LAUGH] >> So if someone were, some young whipper snapper wants to go out and start creating some kind of systematic strategy, is there any alpha left out there? Or where should somebody like that start? >> Oh, I'm sure there's all kinds of alpha to be had. It's just you have to, don't look in the same old places because everybody already has been there and there's only a bit. I think there is all kinds of opportunity if you find new information sources. You hear great stories about little creative ideas where you pay someone sitting in China to count trucks coming out of a factory. >> Right, right. >> Right, you've heard this one. And then you get a sense of how production is going for that company. Very simple little information advantage. >> Right. Mechanical Turk like sorts of things. >> Yeah, yeah, yeah. There's all kinds of things you can do with that. >> So I think that's fascinating. The one thing I would say to sort of reinforce the kind of thing that Quandl can offer is we often find that the combination of different data sources adds value. So as an example, maybe over simplified. Back in the 90s, a strategy, say based on Bollinger Bands, might have worked when you hit the top that could be a cell signal. Nowadays, that just by itself, it doesn't work. [LAUGH] But, a combination of multiple technical indicators plus some fundamental data, putting all those together instead of just a simple, single factor model can work. >> Yeah, I know, I think so. I still have, there's three rules of thumb for building quantitative strategies these days. I think one should look for theoretically sound ideas. So don't bet on black box, neural net, kind of things, because they always fall down, right? But if something makes theoretical sense in an economics or financial sense, it's a good chance you're onto something. >> So it's a good intuitive story to start with. >> Exactly, it actually should make sense on a whiteboard to start with, right? Grounded in some of the fundamental principals that we know govern financial markets in economics right? Number two is empirically tested, right? So build something that's theoretically sound and then show yourself empirically, with data, that this thing worked. And then the final thing is be scared of complexity. If it doesn't stay simple its a warning sign your probably over fitting and getting yourself into a trap. But if you devise a strategy that ticks these three boxes, that it's sort of theoretically sound, empirically testable, and still simple, you're probably on the right track. >> Yeah. [LAUGH] That's great. I'm in total agreement with you. So, one last question. So we've talked about things to do or how to attack it. What should people run from, [LAUGH] what should they be scared of? >> Look I think the single biggest trap, and I've fallen into this thing myself, right, is if you find yourself calibrating the same model on the same data usually, and you calibrate it. Oh, it doesn't quite work. And then you turn a few knobs and it doesn't quite work. And then the eighth time you finally turn the knobs just right and it's working? Odds are you've just curve fitted, right?. Don't fall into that trap of endlessly iterating on the same data. Because you're in for unpleasant surprises once you get into, out of sample data. >> Great. Yeah. [LAUGH] You're totally going to reinforce the things I say in the class, so that's awesome, but they don't believe it when they hear it just from me. >> Trust me. I've lost a lot of my [LAUGH] >> Cool. Well that's it. Fantastic, great interview. Thanks so much for your time Tammer. >> The pleasure was mine Tucker, thank you. >> And hope you all enjoyed it and I'll see you online. So long.