Big Data: Exploiting an Inefficient Market

Disruptive Forces in Investing

July 9, 2019

Anu Rajakumar: Welcome to Disruptive Forces: A podcast from Neuberger Berman where we discuss insights and innovations in the investment management industry. I'm your host, Anu Rajakumar, and in today's episode, we'll hear from Michael Recce, Neuberger Berman's first Chief Data Scientist whose big data team analyzes large, unstructured novel data to evaluate the health of businesses. Michael, welcome to the Disruptive Forces Podcast. Thanks for taking some time to join me.

Michael Recce: Great to be here. Thank you for inviting me.

Anu: Thank you. We're hearing a lot about big data, alternative data sets, machine learning and artificial intelligence. In the context of the investment management industry, could you briefly explain the jargon for us and share some examples of how these are being integrated now into the investment research process?

Michael: Super. Yeah, thanks. So the easiest way to imagine the pieces coming together is to start with where I came from before coming into the industry and that is in ad tech. So, if you went to a webpage in 2009, you would see the same ads as anyone else visiting the webpage. They were like a newspaper of static ads. By 2014, the majority of ads you see are dynamic. They're personalized to you. And this process has led to almost an uncanny set of ads when we look at webpages now—we almost are uncomfortable at how accurate the ads are. Now, if you know what ad to show someone, you know what product they're interested in. If you know that across all products, services, geographies, you know which companies are winning in the marketplace. And it turns out, that's not priced into the stock market. So a simple way of thinking about these pieces coming together, this cloud computing and machine learning and big data is to think about Facebook. So if you get the ideal job at Facebook, that's what you're doing. You're doing machine learning in the cloud on this clickstream messy data in order to figure out what ad to show someone. So in fact I recruit those people. I say look, let's say you get that best job, you should come to finance. If you go into finance, then you get to do the same analysis of the same data, but the positions are ten orders of magnitude bigger rather than just buying an ad. And not only that, you get paid more. So this field of using big data and finance is very much like the field of targeted advertising. It's analyzing people's behavior and all this messy data that's online, using machine learning methods in order to understand what products people are interested in. And it's the same for consumer as it is for cross industry.

Anu: And I imagine there are tons of new sources of data that you could potentially be using for insights. How do you evaluate whether a source is as powerful as you might expect it to be?

Michael: Yeah, that's a great question. It turns out that most of the data that you might want to use isn't useful. In fact, it's probably only about five percent of the data sources are actually useful. And it's a very big task for the team to determine the usefulness of data. The other thing I'll say is that it's not really the data in its raw form that's most useful. It's what you do when you're processing the data. And so there's a bunch of questions we ask and then some evaluations we do with data once we get it in order to figure out its utility.

Anu: Sure. I imagine that a lot of this data is newish in the last few years it's been created. Is there concern about the reliability of vat testing such relatively short amounts of information that you're looking at?

Michael: Yeah. So I think from the quantitative investing point of view, quants like to look at many, many years of history of data so they can so-call back test it so they can determine if a signal is true backwards in time. I mean, it's interesting that in this new field, it's sort of taking aspects of quant and aspects of fundamental. It's interesting to compare these two groups because you'll hire a fundamental investor who can predict the future value of a company. So they're looking forward in time at where this company is going to go and where the managements going to take them. When quants, they're looking at historic things, backwards in time. In fact, they really put forecasting the future completely out of their domain. And so these data are allowing us to sort of think about this combination of the two—of how you would use some of those methods. And so it's less about historically what kinds of mispricing's have occurred in time like quants look at, and more about who's winning market share. And so if I look in real time at all the different products and services made by a company, I can look at which ones are gaining or losing market share with each segment of the population and with each marketplace.

Anu: Now, you mentioned fundamental investors, quantitative investors, I think lots of asset managers are making significant investments in the technology and the data. I've previously heard you say not all data is equally valuable. Could you explain that a little bit?

Michael: Sure. So the simplest example is if you look at the consumers. When the weather's nice, like it is in New York today, people go shopping. So I can look by region and say, you know, where are companies located, and has the weather been nice? And if there's been nicer weather, then there will have been more shopping. Well, slightly better than that is cell phone location. So 20 companies will sell you cell phone location. So if I have cell phone location, I actually know at least you took your phone shopping. I don't know if you actually bought anything but it's better than the weather. And in fact we actually have credit card transactions. So if you have 10 percent of U.S. credit card transactions, you actually know what people bought so it's more than just going shopping. You know what's actually being spent at each store. So the closer you get to the details of the data, as I say, to say to the coal face, the more information you actually have about the impact it's going to have on commerce. So, you might want to look at the other data sets because it'll give you some ability to determine if there's bias in your data, but you want the data which actually speaks in most detail about what's actually happening.

Anu: Right. So separating the valuable information from the noise sounds like it's a pretty arduous task.

Michael: Yeah, I mean, all data sets are biased, and the most important thing is to remove the bias. The second best is to actually detect when you have bias so you know what your conviction is.

Anu: And it sounds like there's a very compelling case about the opportunities to gain from the information edge by evaluating these big data sets. What are some of the challenges or the risks that you can see?

Michael: The challenge is the technology in traditional bricks and mortars businesses, just the wrong technology. Internet companies have developed an entirely new technology stack for being able to process very large messy data sets. Whereas lots of bricks and mortar businesses try to be very conservative and they're sort of 10 years and 15 years old in their use of technology. So I think as firms try to think about how to use data, one of the biggest challenges is rethinking the way they do technology and the way they build that technology organization. I think, all of these data sets have challenges and, Gartner has this process which they call the hype cycle. The hype cycle has an initial very high sharp peak and then it has a deep trough and then it goes back up to a high plateau. And so in this very first peak is more heat than light. People are all trying it. They're all trying to do it. And people get burned and then that's what causes the trough. And then eventually the field comes back up to being used. And I think we're near the peak of this initial hype cycle. And so many firms are trying to get involved in this data and they're going to find challenges that they hadn't previously expected and find it more difficult going than they might have imagined.

Anu: Right. Because I imagine one of the concerns about using big data is that concern about sort of future overcrowding or an alpha decay. How do managers maintain an edge with that in mind and will there be sort of an eventual commoditization or democratization of data in the future?

Michael: Well, I think people tend to worry about this thing called alpha decay. But I think there is a little bit of alpha in the raw data itself but most of the alpha is in how you use the data. I mean, I think lots of quants have been using various data sources for a while and the typical quant process is to remove the outliers, you so call normalize it. Then you shred it and put it in the soup. And so you end up with this big soup of data and you try to find some alpha. Now, what we're doing is completely different. We're actually trying to use data sets which actually tell you something unique and different about the value of the company. They provide you a higher resolution microscope to understand the company. And that type of information, the alpha doesn't go away. In fact, what you can do is you can look at the market and ask how much alpha is left in the market? And it turns out it's enormous. I mean, I came from the technology industry into the investing industry. So I have that advantage and disadvantage. But if you look, one of the first things I heard about was earnings surprise. Now, people talk about efficient markets but if every quarter we're surprised by all the earnings, how could the market have everything priced in? Because we're surprised. And so the thing is that, if you just use earnings surprise and if you add up how much earnings surprise there is in any sector and any quarter, that's a measure of information that's not in the market. And until all that information goes away, there's still alpha in the data.

Anu: So maybe to that end, can you share an interesting alternative data set or a non-conventional use of alternative data set that you found to be empirically useful?

Michael: Yeah, so I think people think that the initial application of data in investing was around use of data for earnings events. But let me just pause for a moment to try to explain this a little bit more carefully. So there are at least three different time domains for thinking about using data. So in one time domain, the shortest time domain, social media, and news and sentiment are used. And so if I know before everyone else about some news item, then it's sort of a race towards high-frequency trading. The first person who can trade on it gets ahead of everyone else. That type of data's been used in the quant space for a while. The second type of data is around earnings events. So when companies tell you about their earnings, if you can predict how the companies doing before the company tells you there are these big earning surprises you can trade on them. And so that was my first experience in finance was trading around earnings surprises. But actually the CEO doesn't tell you very much. He only tells you about top-line earnings. He gives you some guidance. The third time domain which is where we live in now is you can think of it as like longer term investing where you want to pick the next Google before everyone else knows they even exist. And so that requires you to know even more about the details of the data and in that time domain you want information about individual cohorts and demographics and how people are becoming engaged with data. But again, the initial application in these later time domains has been in consumer, but these same data sets are actually useful for lots of other sectors. One example is the use of credit card data for predicting U.S. oil production. So gee, how could I use consumer data around credit cards and bank accounts to predict U.S oil production? Because we actually see the deposits into these accounts as well as the withdrawals. And so you can look at the deposits. And it turns out, about 20 to 50,000 people in the U.S. lease their land for fracking. And there are about a handful of trusts who have managed royalty payments for that land usage. And the royalty payments of course are proportional to oil extracted. So you find a few hundred accounts that actually have these royalty payments. You measure the amount of their royalty. You regress it against U.S. oil production and that gives you a read on a daily basis how much oil is being produced through fracking in the U.S. That's just one example but there are many, many more creative examples of how these data actually explain which companies are winning and losing across all the different marketplaces you might be interested in.

Anu: Yeah, that's a very fascinating and interesting example. Thank you for sharing. So maybe as we think about big data from a geographic perspective. How does the data from developed economies compare to data from developing or emerging markets?

Michael: So this phenomena of actually using data for investing started in restaurants and apparel and started in North American equities. And like lots of things, social media search and all these things, they have a place where they start and then gradually spread to other asset classes in our case and to other geographies. I think we're still in the very early stages of this process and we are currently in central research here building out portfolios for global equities and also my team actually helps teams in other asset classes. Because I think that the opportunity is even bigger in some of these areas which haven't really leveraged these novel data sources at all.

Anu: Michael, thank you so much for being on today's show and sharing your insights on big data. I'd love to have you back again to delve into how big data's changing other areas of the investment industry.

Michael: Thank you.

This podcast includes general market commentary, general investment education and general information about Neuberger Berman. It is provided for informational purposes only and nothing herein constitutes investment, legal, accounting or tax advice, or a recommendation to buy, sell or hold a security. This communication is not directed at any investor or category of investors and should not be regarded as investment advice or a suggestion to engage in or refrain from any investment-related course of action. Investment decisions should be made based on an investor's individual objectives and circumstances and in consultation with his or her advisors. Information is obtained from sources deemed reliable, but there is no representation or warranty as to its accuracy, completeness or reliability. All information is current as of the date of recording and is subject to change without notice. Any views or opinions expressed may not reflect those of the firm as a whole. This material may include estimates, outlooks, projections and other “forward-looking statements.” Due to a variety of factors, actual events or market behavior may differ significantly from any views expressed. Neuberger Berman products and services may not be available in all jurisdictions or to all client types. The use of tools cannot guarantee performance. Diversification does not guarantee profit or protect against loss in declining markets. Investing entails risks including the possible loss of principal. Investments in hedge funds and private equity are speculative and involve a higher degree of risk than more traditional investments. Investments in hedge funds and private equity are intended for sophisticated investors only. Indexes are unmanaged and are not available for direct investment. **Past performance is no guarantee of future results.**

The views expressed herein may include those of the Neuberger Berman Multi-Asset Class (MAC) team and Neuberger Berman’s Asset Allocation Committee (AAC). The views of the MAC team and the AAC may not reflect the views of the firm as a whole, and Neuberger Berman advisers and portfolio managers may take contrary positions to the views of the MAC team and the. The MAC team and the AAC views do not constitute a prediction or projection of future events or future market behavior.

Discussions of any specific sectors and companies are for informational purposes only. This material is not intended as a formal research report and should not be relied upon as a basis for making an investment decision. The firm, its employees and advisory accounts may hold positions of any companies discussed. It should not be assumed that any investments in securities, companies, sectors or markets identified and described were or will be profitable. Any discussion of environmental, social and governance (ESG) factor and ratings are for informational purposes only and should not be relied upon as a basis for making an investment decision. ESG factors are one of many factors that may be considered when making investment decisions.

This material is being issued on a limited basis through various global subsidiaries and affiliates of Neuberger Berman Group LLC. Please visit www.nb.com/disclosure-global-communications for the specific entities and jurisdictional limitations and restrictions.

The “Neuberger Berman” name and logo are registered service marks of Neuberger Berman Group LLC.

© 2019 Neuberger Berman Group LLC. All rights reserved.