ilLUminate Blog Transcript: **Sterling Yan on Human vs. AI Investing Strategies**

Recorded February 1, 2023. Listen to it here.

ANNOUNCER: 00:02	[music] This podcast is brought to you by ilLUminate, the Lehigh business blog. To learn more, please visit us at business.lehigh.edu/news. [music]
JACK CROFT: 00:13	Welcome. I'm Jack Croft host of the ilLUminate podcast for Lehigh University's College of Business. Today is February 1st, 2023. And we're talking with Sterling Yan about whether machine learning methods can actually boost investment performance over results obtained through traditional models. Dr. Yan holds the Joseph R. Perella and Amy M. Perella Chair in Finance in Lehigh's College of Business. His main research interests include asset pricing, institutional investors, mutual funds, hedge funds, short selling, and liquidity. Welcome to the ilLUminate podcast, Sterling.
STERLING YAN: 00:53	Thank you, Jack. Thank you for having me. Thank you for the introduction.
CROFT: 00:57	Before we get into the details of your research, let's start by looking at some of the differences between human investing strategies and machine learning strategies. As you noted during a presentation on your research last year, the entire hedge fund industry is built on the idea that fund managers are able to predict stock returns. So how do successful human investors arrive at their determination whether one stock will have higher returns than another stock?
YAN: 01:28	You're right. Return prediction is extremely important not only among finance academics, but also in the financial industry. As you can imagine, if you're able to predict returns with a lot of success, essentially, you are able to-- you are able to print money. So that creates a lot of incentives for the smartest people to try to predict returns, and hedge funds, the industry is built on the idea that they are able to predict returns, right? They claim to be able to predict returns. Now, even though we talk about human investing strategies versus machine learning strategies in your question, but I don't want to put them in two completely distinct categories. Because even in the machine learning strategies, there is human input. There's a lot of human input in that process. Having said that, traditionally, fund managers, institutional investors use, broadly speaking, fundamental analysis and technical analysis, OK? So fundamental analysis, they would use financial statement information about the company. They use economic analysis, analysis of the industry that the company is operating in. Anything that's relevant for the value of the company. So what they do is they try to arrive at a value of the company, and then compare that to the market price load or stock of the company and try to determine whether they should buy or sell the stock. So that's fundamental analysis. Technical analysis is where fund managers are trying to use historical trading information, historical prices, historical trading value. So by and large, those are the two primary approaches used by fund managers in the past.
CROFT: 03:22	And with machine learning, what information and how do they arrive at the stock return results that they're predicting?
YAN: 03:32	So, with the traditional approach, fund managers tend to focus on a relatively small set of variables or signals that they can use to predict stock returns, for example. And their ability to account for the complex interactions or relations between those

	signals and variables with future stock returns is kind of limited. And machine learning strategy, the new technology allows the fund managers to be able to expand, for example, the set of management signals that they can consider—increase that steps substantially, and be able to identify the signals that truly can forecast future stock returns. Again, that's the theory. That's the idea. That's what they're trying to do.
CROFT: 04:30	And how prevalent has machine learning become in predicting investment performance?
YAN: 04:35	I think it is very difficult to get a precise idea on this, but if you are a large asset manager, if you are an asset manager with a decent size, I don't think you can afford not to look at machine learning nowadays. Having said that, most of them have explored machine learning in their investment process. The extent to which they actually use machine learning in their investment process is an open question.
CROFT: 05:04	So is it more of a case, particularly where you have investment and fund managers use it as a supplement to what they do, as opposed to relying on it to make decisions?
YAN: 05:21	That's a very good way to characterize the role of machine learning in the investment process nowadays. So those are two categories. Some of the fund managers perhaps develop a fund, entirely do it on the idea that the maximum process is going to be done with machine learning strategies. But most of the funds, traditionally, are using machine learning strategies as a supplement to help with their existing investment strategies and methodologies.
CROFT: 05:58	Now there have been a series of studies over recent years, asset pricing studies, that usually find that machine learning can double or even triple stock market returns delivered by traditional models. What was it that led you and your three academic colleagues from other universities to examine whether that was actually the case?
YAN: 06:22	So some of the scholars in recent studies using machine learning strategies to predict future stock returns have, by and large, painted a rather rosy picture for the performance of machine learning strategies. So me and my colleagues-- and there are two broad reasons. One is sort of a general reason why we are a little skeptical about this finding and the other is a specific reason. So let me talk about the general reason first. There's actually a theory in finance and that is the market is pretty efficient. Now the market is not always efficient, but the market is pretty efficient. That's one default theory of the financial market. And it's built on the idea that if the market is not efficient-- when the market is efficient, by the way, stock returns are not very predictable except that one is predictable because they are risky. And the idea why the market may be quite efficient is built on the idea that if the market is not efficient, then smart investors will try to exploit that and the simple process of exploiting market predictability is going to make the market more efficient. So that's why there is the default theory saying that the market should be pretty efficient. So that's one general broad conceptual reason why one can be skeptical about a very large magnitude of predictability documented by previous studies, for example. A specific reason has to do with the design of the studies, in these recent studies, a particular design choice. And it turns out what happened was that most of the recent studies use anomaly variables, investment signals, that is that word discovered in the more recent time period, and they assume that real-time investors in the 1960s, many decades ago, were able to be aware of those predictors and use them to be able to predict returns. So there's a hindsight bias in some sense in that choice of

methodology. I can elaborate on that if you want to, but there is a hindsight bias in some sense that is--

CROFT: 08:45

Right. Yeah. I do think that's interesting. And I think it would be interesting for our listeners to understand that a little better as well. This idea that in looking at these anomaly variables that have been developed, say, from the '90s onward and then applying them back to the '50s or '60s or '70s that there's an underlying assumption that investors at that time in the '50s and '60s and '70s were aware of these variables that weren't actually identified until decades later.

YAN: 09:24

Exactly. I think, Jack, you made it more clear than I did just a couple of minutes ago on that point. So, again, this has to do with the field of finance being different from the other fields, for example, where machine learning and artificial intelligence have been very, very successful. For example, I mean, artificial intelligence can play chess, right, much, much better than human beings are able to. And there are many, many other areas, image recognition, cars driving themselves. There are many, many other areas where artificial intelligence is extremely successful. But there are some fundamental differences between finance, between return prediction, trying to predict future returns from the other applications like I just mentioned. Meaning that in finance, in order for machine learning, artificial intelligence to be successful, one of the things that's necessary is that you ought to have abundant amount of data, a large amount of data for the machine learning, for the artificial intelligence to be able to learn from. That's actually not the case in finance. We can't artificially generate new data, right, come up with images or let a car drive itself or let the artificial intelligence to play chess games to create millions and millions of new chess games. In finance, we can't generate data that way. The market goes down today by 1% or it goes up at 1%. That's the data we have. We can’t artificially generate a new set of data where today, the stock performed differently. So the amount of data is relatively limited in the finance area. And there are some other differences where finance is different, and that is one of the reasons why machine learning strategies may not be as successful as one would expect.

CROFT: 11:25

Now it does seem-- the stock market has been around for basically a century or more. So there is what would seem to at least a lay person, there must be a lot of information out there. But I know in your study, you addressed that 100 years really isn't that much for what you're talking about.

YAN: 11:46

You're right. Your question is extremely reasonable to a lay person. Now to an expert, 100 years of data where trying to-- there's a lot of-- that's the second characteristic that I haven't touched on and that is there's a lot of noise in finance data. When I say noise, meaning that if you’re trying to drive a car around based on the information regarding speed and distance, for example, you can learn and improve your driving skill and that's very-- because the data you receive, the feedback received is extremely informative. In finance, however, because the data is very noisy and because most of the day, stock prices go up and down because of new information or because of noise, not because the predictable component of the return that allows you to say the market will go up or go down. And as a result, it takes a lot more data, it takes a lot longer time series in order to learn the rules, okay? You're not learning something that's-- in driving a car, for example, you are trying to learn some physical laws, right, the relation between distance and speed, how you make turns, how you change directions, how that's going to affect your safety, right? And here, you are trying to learn something where the data tells you very little. There's a lot of noise. That's why it takes a lot more data to learn where the data is actually relatively scarce. That's the reason why is that 100 years of data is actually not very much.

YAN: 13:31

And another characteristic, I think I will put this point on right here at this point as well. So I've talked about there's relatively little data in finance and there's a lot of noise in financial data. And the third characteristic that makes machine learning in return prediction extremely difficult is that financial markets are adaptive. We're not trying to learn something that's fixed, constant. That's actually adaptive. It's going to change the moment you are trying to learn. And that goes back to one of the arguments I made earlier. The very process smart investors trying to explore in any predictability in the stock market will make that predictability go away. I don't know if that's intuitive, but that's the idea. If something predicts the stock return, if a lot of smart investors is trying to use that investment signal to predict stock returns, it's trying to implement that strategy that will make that predictability actually disappear, which means you have to relearn. And that's something that's also different that distinguishes the financial market from the other areas where machine learning is successful. It makes it more challenging.

CROFT: 14:45

Now as simply as you can, if you could talk about the methodology you used in your study. And then I think what most people are most interested in, of course, is kind of the bottom line. What were the main findings of your study?

YAN: 15:02

Previous studies basically use those anomaly variables that are discovered exposed as predictors as if the investors were aware of those in the 1950s or 1960s, right? And that was the flaw we identified over the-- of their methodology identified. Again, the reason why they did that has to do with the limited financial data that we talked about. We can't generate new data to train the-- to train the model, we actually have to use past data in order to test whether the model works or not. So there are reasons for why they did what they did, but that was not ideal. That was not perfect. So what we end up doing was let's take a real investor's perspective. Let's assume we're in the 1950s and '60s. We didn't know what was going to happen in the 1980s, right? We didn't have the hindsight. And therefore, we're going to do is we're going to look at the data we have. We're going to construct a bunch of strategies. We're going to call that a universe, OK? We didn't know in the '50s, '60s which one was going to work exposed. So what we're going to do is we were going to learn from the universe. So what we end up doing is we tried to simulate that process. We construct a universe of signals over 18,000 of them. And assuming the investors were able to use machine learning strategies in the '50s and '60s and try to learn something from that universe, and then use that to enhance our implemented investors strategy. And then we evaluate the performance of that strategy. So what our main findings are that yes, we find that machine learning strategies do work in the sense that they can enhance your investment performance, but not to the extent that was documented in the recent studies that we just talked about. So the magnitude of the investment performance improvement is substantially smaller than what has been documented and again, for reasons of the limitations of their methodology.

CROFT: 17:06

If I understand this correctly, rather than just stopping with pointing out the flaws that you were able to identify in the previous studies, you and your colleagues went on to develop real-time machine learning investment strategies with the key being

	that they're implementable in real time. So if you could talk about how you did that and why that's so important.
YAN: 17:31	So essentially, we just talked about the construction of the universe of strategies and not assuming investors were aware of what was going to happen in 20 or 30 years later. And that made our strategy real-time implementable. And that's important because when we evaluate the performance of the investment strategy based on machine learning methodologies, we want to make it to be real. So that's why it is important, and that's the key difference. In fact, the title of our paper is the Real Time Performance of Machine Learning Strategies. So that dimension is important to our study.
CROFT: 18:15	And so what's the bottom line-- let's start with the bottom line of your study for those in the investment field. What kind of takeaways are there for them if-- and I guess there really aren't any who are not using machine learning to some degree at this point, but how do they change what they're doing then?
YAN: 18:41	Machine learning is changing everything. Machine learning holds a lot of promise for the financial industry as well. Having said that, return prediction is fundamentally different, more challenging than some of the applications where machine learning, artificial intelligence has been very successful at for the reasons I just talked about. Financial markets, we have limited data. Financial data, not as high quality because the information ratio-- the noise ratio is too high. Information ratio is too low. And financial markets are adaptive. They change. It's dynamic. You are not learning something that's fixed. So for those reasons, machine learning strategies is going to face a lot more challenges in the return prediction in the financial industry than some of the other areas. So that's something we want the investors and fund managers to be aware of, to be more realistic about what machine learning strategies are able to help them with their investment strategies.
CROFT: 19:51	And in terms of investors, and I would assume that the people relying more heavily on machine learning are probably the smaller investors who, because of the cost differential and all that, have opted into that. What should they be thinking about in light of what you found?
YAN: 20:12	Well, at this point, machine learning strategies have become more readily available even to some of the smaller investors, but I would argue that this is still something where large investors-- actually, managers with a lot of resources are more capable of doing at this point. So for smaller investors, especially individual investors, this is not something that has become a mainstay among the retail investors.
CROFT: 20:44	And in terms of kind of looking at a crystal ball into the future, as machine learning continues to grow more sophisticated, how do you think the role of the human fund managers will change in the future?
YAN: 21:02	That's a very good question because we can never underestimate how fast or how much the technology is able to change, right? So machine learning strategies themselves will become more sophisticated, perhaps will be reaching at a stage where it can do things that are not thought possible now today, even in the area of finance. So that's possible. But those fundamental characteristics of financial markets that I have talked about maybe two or three times already in this podcast, those things will not change very quickly. And as a result, the potential of machine learning strategies in the area of return prediction, in my opinion, will continue to not be as

impressive as what we would expect or we would have hoped. And because of that, I think human investors, human managers will continue to play an important role in the investment process. And in the future, maybe some sort of a combination between human beings and artificial intelligence. That may be the optimal model going forward. The machine learning will likely to play an increasingly important role, but there will not be any substitute for human involvement in this process.

CROFT: 22:29 And the last question I usually like to ask is, is there anything I haven't asked you about or anything that we haven't talked about that you think our listeners should know about machine learning and stock returns?

YAN: 22:43 I just want to reiterate that, again, in general, artificial intelligence, machine learning hold considerable promise in many different areas of our life or career in different areas. Finance is just somewhat different, especially in terms of return prediction. There are some areas in finance where there's abundance of data where the goal is different. It's not necessarily return prediction, in which we have already seen a lot of success of machine learning. That's not to discount the value of machine learning in finance in general. In this specific area, where you are trying to predict the future, predicting future stock returns, where it proves to be very challenging even for the machine learning strategies, artificial intelligence with today's technology. So that's probably my last word on it.

CROFT: 23:35 Okay. Sterling, I like to thank you again for joining us on the ilLUminate podcast today.

YAN: 23:42 Thank you very much, Jack, for having me. I hope that discussion is helpful.

CROFT: 23:47 I believe it has been. Sterling Yan has published extensively in top academic journals, including Journal of Finance, Journal of Financial Economics, Review of Financial Studies, Management Science, Journal of Financial and Quantitative Analysis, and the Journal of Accounting Research. His research has been cited in Business Week, The Economist, Financial Times, and The Wall Street Journal. This podcast is brought to you by ilLUminate, the Lehigh Business blog. To hear more podcasts featuring Lehigh business thought leaders, please visit us at business.lehigh.edu/news. And don't forget to follow us on Twitter @LehighBusiness. I'm Jack Croft, host of the ilLUminate podcast. Thanks for listening. [music]