The Art and Science of Big Data – A Quantitative Manager's Perspective

Conversation Series…

Elena Khoziaeva, CFA, Bridgeway Capital Management

Mehrzad Mahdavi, PhD, Executive Director, FDP Institute

Summary

Digital transformation, with artificial intelligence and machine learning technologies at its core, has created unprecedented challenges and opportunities across all industries. The financial sector will be profoundly affected as demonstrated by numerous use cases including asset allocation, security selection and risk management.

As an example, asset managers using artificial intelligence and machine learning, are found to be growing their revenue 1.5 times more quickly than the rest of financial services. Asset managers use alternative data to gain an "edge" over competitors that use traditional or "old" data sources such as quarterly corporate earnings or low-frequency macroeconomic data. Going forward, systematic strategies deployed by quants such as alternative risk premia, trend following, and equity long/short will increasingly adopt machine learning tools and methods along with alternative datasets. This trend is driven by smaller funds that leverage recent advances in artificial intelligence (AI) and machine learning, the explosion in data availability and inexpensive cloud computing to generate alpha at lower costs.

Recently, FDP Executive Director Mehrzad Mahdavi interviewed Elena Khoziaeva, CFA, portfolio manager at Bridgeway, to get her views on systematic investing and use of AI and machine learning in portfolio management. This interview: "The Art and Science of Big Data" brings focus to opportunities and challenges for financial firms adopting AI and machine learning technologies.

Many published articles by FDP Institute, AIMA, CAIA, CFA and others speak of the disruptive forces of AI/Data Science in the industry; therefore, we will assume readers can reference those for background. For this interview, I have curated the most frequently asked questions and Elena's responses. The common thread is that adoption of AI and machine learning in finance will require massive education (and re-education) in Science, Technology, Engineering and Mathematics (STEM). Elena also reminds us of the important role that financial professionals play in this team effort. She brings forward the value of being creative, asking questions, and remaining true to the "art" of data science.

Mahdavi: Please tell us a little bit about you and what you do at Bridgeway?

Khoziaeva: I carry several hats: I on the Investment Management Leadership team together with our CIO, Head of International Equities and Head of Research. Overall, I am responsible for U.S. Equities portfolios. I am on the Portfolio Management Team together with several other portfolio managers. In addition, I am on the Research Team together with other researchers. So, I am on several teams at Bridgeway. Teams work together, teams share knowledge, teams make better, more informed decisions than individuals. Please keep this team approach in mind as we talk about the human aspect in investment research and modeling.

Mahdavi: Let's talk about the nature of being a quant firm, what defines Bridgeway as a quantitative investment manager?

Khoziaeva: Quants are systematic investors. Quantitative investment management firms run statistical and factor models, they optimize around return requirements and risk parameters, quants run simulations and long backtest going back many years; they act on historical information that they have analyzed and created predicting signals. At Bridgeway, we call ourselves a statistically-driven evidence-based firm. What does that mean: we rely on historical data in our decision-making, we run factor models for our stock selection, we optimize portfolios based on statistical results. The key to understanding the quantitative process is that is systematic. Systematic vs. discretionary. Systematic investing is a data-driven repeatable approach that relies on statistical analysis to identify a set of names from a large universe of stocks. That's science part of the process. But at the same time, we absolutely use human skills and expertise to put it all together. It is crucial to consider economic theory. It is absolutely crucial to use common sense. We, humans, we are the best in one important area – that is asking questions. We do ask a lot of questions, most of which would make you think that we are toddlers. They start with "why". A very simple question if you think about it…but a crucial one. Why would you expect this particular factor to perform in this market environment? Why not? Why the result in this this period of time be different from another period of time? Why would the holding period make a difference? When turnover is so high? Why do you see decay? Why do you see increased correlation? And here comes the most important "why" of them all – why would you believe that this historical signal can persist in the future and lead to long-term outperformance? So that's what we do – we run numbers and we ask questions.

Mahdavi: You use the term "systematic" investing. Are all systematic investors alike or are there differences?

Khoziaeva: All systematic investors follow the same philosophy. But there is quite a spectrum there. We all start with the same objective – how can we design a process, how can we form a theory, test the theory, validate the theory and then implement it in practice. Systematic investors believe that designing rules upfront and then implementing them with a minimal touch allow them to overcome various behavioral biases and emotional decisions that tend to lead to poor investment results. Early adaptors of systematic or quant investment, followed the same philosophy as do the professionals now, they just did not have all the computer power that is available currently.

A newer approach involves firms focusing on building databases, using data and big data and computer systems that can effectively perform a specific task eventually without a human intervention, relying on various patterns and inference instead. So, the goal is to solve the problem by creating a sequence of actions that a machine can be trained to do early on and then implement it. Some firms are successful at that; some are not as successful at designing those models; many are combining various signals to create a "super" composite signal. I would say that we are somewhere on that spectrum and as we talk about development of Bridgeway and the growth of Bridgeway, I will definitely touch upon the changes. You asked how machine learning and big data are coming in – that's what is developing, the new growth is in the area of implementing more and more computer power, making processes more efficient, faster and preserving space for humans to use their creativity for the "art" part of the investment process. That is the path leading to the future and I believe that Bridgeway is on that path of using computer power and machine's intelligence to help humans be more creative and innovative individuals on that journey.

Mahdavi: I noticed that data is a big part of your investment philosophy, how do you define big data and alternative data?

Khoziaeva: First question was, how is data used for a quantitative manager like us? I would throw one word that may confuse you. That word is "fundamental". We use fundamental data that comes from the financial statements. We include items from balance sheets, income statements, cash flow statements, we look at most recent financials and look back in time. We also use price data, which is a part of technical analysis. We calculate ratios and variables and we look for patterns. So, we use fundamental data, we use technical data, how can we state that we are quants? Because we process it in a systematic way. Those are the traditional data sources, we started using them years ago, we still use them.

Another part of the question, what is big data? large volume of data – both structured and unstructured. But it is not just the amount. The complexity of big data is driven by the 4 V's: volume, velocity, veracity, variety. All of these dimensions make the big data complex, challenging to process but at the same time it creates an inspiration to look for solutions.

And finally, what is alternative data? that comes from non-traditional sources, not financial statements like social media, companies' commentaries, earnings calls scripts, video images, companies' reviews.

Mahdavi: Tell us about your journey, where you are, where you are going in terms of data driven decisions and what are the differences in terms of people, process and technology?

Khoziaeva: Before I cover data, let me mention that the most important capability that helps companies succeed is developing a strong team. People make all the difference in the world. The people that you bring together in the investment management team are responsible for results. It is about the diversity and bringing people with various backgrounds who can then put their expertise together, bring their perspectives and then come up with ideas of dealing with data challenges, process challenges and model

challenges in a collaborate way. So, the first step is to create a strong team by bringing creative individuals who are ready to put their best forward.

Speaking of data: conceptually, majority of our data is still fundamental and price data. We do need long-term sets of data going back in history and alternative data sets are still not as profound and do not provide enough of a sample size. It does not mean that we are not looking into it, but we are still using mostly fundamental data and price data, but we are processing it faster. For example, using visual techniques to see data availability and coverage is a great tool to review data availability historically and notice holes in data coverage.

I do want to mention data complexity and how challenging it is to address it. I want to refer to one document that has been very helpful to us in our ability to manage the data. I think it is great to have a balance between some guidance and standardization of data analytics and at the same time have flexibility so that each firm can develop to tailor the standards to their unique processes, available resources and timelines. What we found rather helpful was the Sound Practice Guidelines for Quantitative Investment Managers, presented by the CQA group at their semi-annual conference. One of my colleagues, attended the conference and then brought the insights back to us (drafts were presented in 2013, final version in 2014). It has been incredibly helpful to us to structure the data governance, as well as model development process. It provides suggestions to the quant firms in the areas of data governance: data catalog, data ownership, requirements, licensing; describes best practices for steps to load, process and store data. There are some critical steps along the way to think of: data checks, checks for missing or stale data, threshold checks for data changes, checks against an independent source, etc. It is helpful to have a guide on how to address this and how to put it all together. You do not have to adhere to that tightly, but it is a helpful roadmap for the firms to be developing their processes.

Similar to the data side, on the model development side, it is important to create and update model catalog (a set of files containing detailed description, usage, validation and change control of each model). Following the creation of the model catalog, a firm should have sound model management protocols, including model validation, change control, documentation and testing.

One of the things that we have implemented was creating a separate testing environment. Years ago, we did not have a separate testing environment. But those risks and challenges can be costly. Currently, model deployment does not happen automatically, it goes through an extensive coding and testing process in the "testing" environment.

Mahdavi: What do you see in the future? Do you see continuing advances in using AI techniques? Where does your firm and the industry headed?

Khoziaeva: Definitely evolving, definitely developing, there is no choice of staying behind. The technology is changing, there are more and more tools and data available. But I would caution that it is very important to be careful and mindful in that process of moving to something new that may have its own issues.

Where are we? I would say we are aware about the development, learning, taking our steps and setting our priorities to make sure we focus on the right and most promising ones.

You mentioned that based on some studies only about 10% of portfolio managers are using AI and ML techniques. I believe this was published by the CFA Society in 2019, called AI Pioneers in Investment Management. It is a great paper, it highlights what has been done so far, what can be done in the future. It does talk about the five major hurdles to successful adoption of AI and big data in investment processes: cost, talent, technology, leadership vision, and time. I do agree with them - each one is valid on its own and it is important to be balancing potential benefits of developing AI techniques with these hurdles that are costly to overcome. All of these items need to be evaluated.

Bridgeway is taking the position that artificial intelligence's best use is to complement our current processes in order to best combine machine power, the efficiencies achieved by machine processing and the science behind the model development, and human's art of asking the right questions, noticing breakdowns, seeing the big picture or rather small suspicious details. We definitely are on that journey. A specific example that I can mention would be that our team member has created a visualization tool that assists in research by providing a visual multicolor image of a factor efficacy in multiple universes. It is fast-forwarding our ability to evaluate factor robustness across universes.

Another application of AI that we see is using natural language processing for alternative data collection and analysis. We have done a project last year where we were able to scrape certain information and put it in the database form. There is more do there; there are now data providers that allow you to evaluate and analyze transcripts from conference calls. These are interesting projects but what we are cautious about is whether AI and ML can be used to develop factor models to improve excess returns.

I do want to caution us against one of the most well-known issue when using AI – overfitting. Data overfitting is a specific challenge that a lot of firms are facing. Overfitting occurs when a model picks up noise instead of signals. Overfit models have very good in-sample performance but fall apart out-of-sample because there is unseen data there. By its nature, investment data has very low signal-to-noise ratios, which means a high error rate in forecasting stock returns. There are definitely techniques that are being developed to address that and to be able to train

an algorithm to at least minimize the overfitting issue. At some point Bridgeway will be looking at them and implementing them. One of them is called "Forecast Combinations", where, depending on the project, you can combine forecasts from different classes of algorithms, forecasts based on different training windows, forecasts that use different factor libraries or different horizons. That makes a lot of sense to us (it is similar to steps in our model development process – combining signals and seeing if they work better together). Another technique is called Feature Engineering which is a more expensive and time-consuming approach, so that may be something for the future. Feature engineering is based on asking a lot of questions. By the way, feature engineering is a great example of science and art working together to improve the outcome. That is where you are helping algorithm to learn, you are training it and setting it up for success. When asking for example what are we trying to forecast, which factors are likely to provide valuable information or what training windows to use, you are actually combining the art and science to improve the outcome. So that may be something in the future for us.

Mahdavi: What kind of team / skill set do you need to deploy AI and machine learning?

Khoziaeva: It is important to build a diverse team. One of the main things I am looking for is diversity, how to complement the skills that we have on the team. Cognitive diversity is the inclusion of people who have different ways of thinking, different viewpoints and different skill sets. Bridgeway is very intentional about hiring best talent in different fields; currently on the team we have professions with background in engineering and biomolecular engineering, accounting and auditing, statistics, physics, and of course programming.

For our future hires in the investment management area: definitely programming skills (our overall goal is to be using Python is the uniform language in research), financial engineering, data science, data management, attention to details and ability to maintain strong documentation.

I am looking for indications of creativity, innovation, not being stagnant in their development, someone who is looking for new skills. I would also say that cultural aspect is tremendously important to us – being collaborative, humble and not being defensing in your work. All of these qualities are important to build a strong performing team.

Mahdavi: Thank you Elena, would you like to provide some closing remarks?

Khoziaeva: I have enjoyed this, I hope it has been encouraging for our participants, but learning about the history and know current development. Again, pursue your designation and start studying early. Thank you for following me on this journey through time, using Bridgeway's evolution as an example.

I am going to close with the quote of one of the greatest minds of all times – Albert Einstein:

"The greatest scientists are artists as well. "

Don't forget that. Do rely on data, respect machines, process numbers, but also please do allow your creativity, intellectual, skills, experience, and even your intuition to be free. Look for that optimal balance between science and art.