UPSCOPE

Does your career plan have a vision for big data?

By Ed McCarthy

The source, who agreed to speak off the record, had a solid professional background: an undergraduate degree in finance, the CFA charter, and portfolio management experience. But when asked if that combination would qualify him for an investment analyst or portfolio manager position at his current hedge fund employer, he candidly responded that it would not.

Credentials like the above might qualify an applicant for a nontechnical role at the firm, but job candidates with backgrounds in STEM—science, technology, engineering, and mathematics—have a much better chance at landing an interview. And though work experience at data-driven firms (e.g., Amazon, Facebook, or Google) would be a positive factor in an applicant's favor, even candidates with those qualifications would have to pass a coding test to advance in the hiring process.

Welcome to the age of big data. Data-driven investment management will become more pervasive. News articles imply that programmers and computers are replacing traditional portfolio managers. Headlines shout, "How Computers Trawl a Sea of Data for Stock Picks" (*Wall Street Journal*) and "A New Breed of Trader on Wall Street: Coders with a PhD" (*New York Times*). Though a relatively small segment of the industry has fully incorporated big data analytics into its work, the ongoing fusion of data science and portfolio management is likely to affect financial analysts' training and skill requirements. Other industries' recent histories—retailing (Amazon) and advertising (Google), for example—indicate potentially significant disruption to the traditional investment management model. Firms that successfully incorporate data analytics earlier rather than later could gain a competitive edge in the perpetual search for alpha.

So, is it time to quit your finance job and reenroll as a full-time data-science student? That's not necessary (unless you want to change career tracks). There is a more appropriate range of responses available that will help traditionally trained financial analysts adapt to the changing environment.

WE'VE BEEN HERE BEFORE

Data science isn't the first disruptive quantitative approach to investing. Readers with a few decades' experience will recall the focus on financial engineering in the late 1980s and early 1990s. Similar to today's data analysts, many of the first financial engineers—the quants—were STEM transplants. (I recall a meeting with the manager of a large bank's derivatives trading department whose analyst staff largely consisted of PhDs from the former Soviet bloc, including genuine rocket scientists. Most knew nothing of finance and spoke limited English when hired, but their training allowed them to work with high-level financial mathematics and write the code required for the resulting pricing models.)

The current transition is from financial derivatives research to more general business problems that aren't necessarily tied to financial mathematics and theory, says Stuart Kozola, computational finance products manager with MathWorks in Natick, Massachusetts. "From

Illustration by Kelly Alder

my perspective, I see data science as kind of the quantitative finance 2.0," he says. "They're tied more into an organization that has some problems. 'Here is the data we have; let's see if we can try and gain some insight by using the different types of data analysis, statistical analysis routines.'"

TWO TRACKS, SAME DESTINATION

Quants did not put traditional analysts out of business; firms incorporate both approaches successfully. But can analysis driven by big data coexist with the traditional approach?

A 2015 white paper from Citi Business Advisory Services ("Big Data & Investment Management") provides an informative example of how the methods can differ. The paper describes a hypothetical case in which a "discretionary fundamental" portfolio manager meets with executives from an energy company she follows. The executives' body language provides cues that they are excited or nervous about the upcoming quarter. She conducts additional research by reviewing transcripts of the executives' recently published remarks and makes site visits to the company's facilities for impressions of business activity. Contacts in her professional network indicate that the company's staff has been traveling overseas to one area more frequently than usual. Combining those inputs with knowledge of the company's undeveloped energy fields leads her to conclude that the company may soon announce the opening of a new production facility.

This is classic analysis: weaving disparate information sources and fact patterns into actionable investment propositions. But the Citi paper points out that a data-driven approach could reach the same conclusion with an automated analysis of structured and unstructured data. The research could start with a linguistic analysis that scans documents, such as interviews and conference call transcripts, to identify executives' increased repetition of specific phrases (e.g., "research and development") and their increased use of positive, forward-looking statements. If the search results indicate sufficient changes, the analytic software could then query public records to identify the company's undeveloped properties, retrieve satellite images of those properties, and determine whether any sites exhibit noticeable changes (such as increased land clearing or building).

The automated research could next shift to social media sources. An online query might find a presentation listing key members of the company's site development team. A search of those persons' social media posts could reveal a preponderance of recent photos with geotags from the same location of one of the company's undeveloped sites. These results could lead the automated process to the same conclusion as the analyst about a pending announcement of a new energy field. In a fully automated process, the software will evaluate investments that incorporate the analysis, place trades, monitor positions, and close positions based on predetermined price actions.

COMBINED SKILLS

A September 2015 Goldman Sachs white paper ("The Data Revolution") highlights how data analytics can benefit the investment process: "Big Data can enable investment managers to see hidden connections and relationships between companies, including across industries. We believe data analytics can enable investment managers to see these hidden relationships faster and sooner than market participants who are not similarly equipped—leading to a potential advantage in selecting investments."

That conclusion and the prior example naturally lead to the question of how an organization should include big data in its workflow. More specifically, should analysts and portfolio managers have data science expertise? And is it possible to combine both skillsets effectively?

The responses to those questions vary. Hedge fund spokespersons report that hedge funds already combine the data analytics and financial analysis roles successfully. For example, an April 2015 *Wall Street Journal* article ("How Computers Trawl a Sea of Data for Stock Picks") provided insight into how New York City–based hedge fund Two Sigma Investments has implemented a fully automated, data-driven process. The firm's investment analysts and managers (mostly STEM graduates, many with advanced degrees) build models that incorporate multiple structured and unstructured data sources to search for correlations that can predict price movements. According to the article, risk management software evaluates each potential investment, and if the idea clears that hurdle, the requisite trades are placed automatically.

A background source at one hedge fund explained that his firm and its competitors (as far as he knew) do not recruit individuals with MBAs or CFA charterholders. The reason is that in these firms' experience, it is easier to teach finance to STEM graduates than it is to teach data analytics and coding to traditionally trained finance professionals.

Marc Greenberg, CFA, managing director, director of research, and head of professional development at Point72 Asset Management in Stamford, Connecticut, agrees that the skillsets can be combined. "[The legacy] was that the data scientist is one type of person, and a traditional, fundamental person—undergraduate, business, something like that—is another type of person," he says. "We are increasingly of the view that need not be the case. They can be the same person. They can have an engineering or math background; they can know how to code and be fascinated by stocks. They're not mutually exclusive." Point72 has incorporated that view in its internal training programs. In addition to hiring data scientists, the firm is teaching its new, non-STEM hires "some basic computer science skills, mostly around statistics and regression, so that they can have a better understanding of how to use data," says Greenberg.

WHETHER A FIRM COMBINES OR SEPARATES THE DATA AND FINANCIAL ANALYSES, THE SOURCES WITH WHOM I SPOKE AGREE ON THE NEED TO INCLUDE A TRADITIONAL FINANCIAL ANALYSIS PERSPECTIVE AT SOME POINT IN THE INVESTMENT PROCESS.

SEPARATE SKILLS

Mark Ainsworth, head of data insights at Schroders Investment Management in London, believes it doesn't make sense to search for analysts with expertise in both data science and traditional finance. "It's just not likely to find such a person; it's a ridiculously broad range," he maintains.

Schroders hired Ainsworth in late 2014 to form a data analytics team. The unit is tasked with three functions: (1) identifying and sourcing alternative data sources, (2) translating the data and analytics into a format the firm's investment analysts and fund managers can use, and (3) integrating the data with the firm's equities management process. The data group shares floor space with the equity managers, and its team members regularly attend investment staff meetings to facilitate that integration, says Ainsworth.

The unit provides both reactive and proactive support. Reactive analyses result from analysts' requests; Ainsworth gives the example of a portfolio manager who was skeptical about a product purchase and use claim from a company about to go public. "[We] looked into some sources to find out what the actual customers of that product described as the reasons why they used it and the likelihood that they would repeat-buy that product," he says. "That suddenly turned what was a very uncertain sort of personal anecdote–based element of their investment thesis into a kind of solidly evidenced piece and influenced their decision."

The proactive support uses data research to seek out potential investment themes for the investment staff. Computers can constantly monitor for potentially useful developments that could lead to further analysis, which in turn could lead to investments or divestments.

Another goal for Ainsworth's unit is to offer its members career options. Those who want to become data scientists can remain on that path, but others who wish to become fund managers can eventually move to that track. The goal is to develop cross-trained analysts and fund managers who are proficient with both analytics and traditional financial analysis, although Ainsworth expects that those moving into fund management will still rely on the data scientists because of the time demands to maintain investment expertise. "But they'll know what sort of questions can be asked [of the data unit] and how they can be answered, which is incredibly valuable," he adds.

AVOIDING THE SPURIOUS

Whether a firm combines or separates the data and financial analyses, the sources with whom I spoke agree on the need to include a traditional financial analysis perspective at some point in the investment process. Dan diBartolomeo, president of Northfield Information Services in Boston, gives a hypothetical case in which data analytics allow a firm to calculate the correlations between 10,000 variables and the S&P 500 Index. Given the broad scope of the analysis, it's likely that multiple variables will correlate well with the index. Although some of the variables will have a logical link to the index results, he cautions that other relationships will be spurious. "If it turned out that something that correlated with the S&P 500 was, say, milk prices in Pakistan, I would probably say not [to include it]," he says. "I think the role of the financial analyst is to understand the fact that these big data analytics don't have judgment. They can find relationships, or they can say, 'When this happens, that happens.' The fact that that's a coincidence or there's some plausible reasoning as to why these two things are related, I think, is very much in the domain of the financial analyst."

RAMPING UP THE SKILLSET

A review of online postings for data scientist jobs shows that larger traditional investment management firms, such as banks and mutual funds, are expanding or at least actively exploring the use of data analytics. Building out a data unit is an expensive approach that midsize and smaller firms are unlikely to adopt; those firms will instead probably work with third-party data analytics providers, which are appearing more frequently in the US and abroad.

RavenPack, headquartered in Málaga, Spain, is an example of the latter group. According to Peter Hafez, RavenPack's chief data scientist, the firm's analytics transform unstructured big data sets, such as traditional news and social media, into structured granular data and indicators to help financial services firms improve their performance. Northfield also offers a news analysis service. "We monitor the text news—in our case, coming through the Dow Jones system," diBartolomeo explains. "It's about 3,000–4,000 articles per day, and we do various forms of analysis, which allows us to update our opinions, particularly about the riskiness of different stocks, sectors, countries, and markets."

Does it make sense, then, for traditional analysts and portfolio managers to learn more about data analytics and programming even if their firms aren't hiring Google alumni? DiBartolomeo believes the ability to distinguish genuinely valid statistical relationships from background noise is more important than coding skills. "I think what's important is that financial analysts have a grounding in proper—I want

to call it 'statistical'—logic," he says. "How do you think through these problems and understand, when you see data and you see it presented in various ways that you perceive to be meaningful or not? I think this is part of the job. It's more about, I think, having the ability to rigorously evaluate the stuff coming out of these processes, because the processes can just bury you."

Catherine Truxillo, senior manager in the advanced analytics education group at statistical software firm SAS in Cary, North Carolina, agrees that it's worth the effort for investment analysts to develop stronger statistical skills. Effective communication between the data and investment groups requires some knowledge overlap, she says. "Data analysts and data scientists need to know a bit about the subjects that they are working on because lacking the knowledge foundation can result in analysis oversights and inefficiency. Similarly, investment managers working with data scientists should be versed in the problem-solving approach and analytical project cycle that the data analyst or data scientist uses so that analytics are appropriately leveraged within the financial organization."

Truxillo also notes that data analytics typically involve multiple participants filling different roles; analytics are rarely a one-person role. "In reality, most data scientists specialize in one area or another of data science because data science is collaborative work," she says. "For example, you may have a data-science team that is composed of individuals who have a unique focus on either data management, mathematical modeling, data mining, or business acumen."

TRAINING OPTIONS

SAS provides multiple opportunities for its users to study data science. The SAS Academy for Data Science offers three programs that lead to professional certifications; other courses and programs cover business analytics and focused themes. Colleges and universities also provide data-science courses and offer certificate programs in business analytics.

Another option is to develop proficiency with what Kozola classifies as second- and third-tier analytic software. The first tier is the underlying database, which is the data scientists' domain. Second-tier software focuses more on customized analytics; he includes languages like MATLAB, Python, and R in this category. Learning to use these languages is still challenging for nonprogrammers, but their integrated development environments are easier to master than languages like C++.

Visualization and presentation applications make up the third tier, and Kozola cites Tableau and Click as examples. These programs connect to the data and provide a variety of different point-and-click tools to build dashboards that allow users to perform simple calculations and visualize their data in different ways.

AMONG TRADITIONAL INVESTMENT MANAGEMENT FIRMS, THE CURRENT THINKING IS HOW BEST TO USE DATA ANALYTICS TO SUPPORT THE INVESTMENT MANAGEMENT STAFF AND HOW BEST TO INTEGRATE THE TWO PROCESSES.

Kozola believes data analytics software will continue evolving to the point where users will have simplified access to tools like machine learning and advanced statistical concepts. Visual composition frameworks could allow users to drag-and-drop the desired analytic tools into a dashboard, possibly eliminating the need to use programming languages. "I can see a lot of business schools, a lot of business analysts, a lot of financial analysts moving to that once it becomes mainstream and well established," he says.

NONSTOP LEARNING

Both the on- and off-the-record sources for this article see two investment management tracks emerging. Hedge funds are the most visible examples of the first track, in which a STEM background and solid coding skills—preferably with data analytic experience—are the minimum qualifications for investment analysts and managers. Among traditional investment management firms, the current thinking is how best to use data analytics to support the investment management staff and how best to integrate the two processes.

Four years or more for an undergraduate degree, two years for an MBA, and three or more years of studying 10–20 hours each week to earn the CFA charter is a lot of work. The good news is that those efforts develop knowledge and skills that will remain in demand. At the same time, investment management is dynamic and data science will continue to spur advances in machine learning and artificial intelligence around the financial markets. Unless you are approaching retirement, resting on your laurels and ignoring that trend is no longer an option—if it ever really was.

Ed McCarthy, a freelance finance writer in Pascoag, Rhode Island, is writing a book on computational finance.

KEEP GOING

"Making Sense of Big Data: Nate Silver on the Signal and the Noise," *Enterprising Investor* (11 August 2015) [blogs.cfainstitute.org/investor]

"Little Big Data," *CFA Institute Magazine* (November/December 2014) [www.cfapubs.org]

"Book Review: Numbersense: How to Use Big Data to Your Advantage," *Enterprising Investor* (5 September 2013) [blogs.cfainstitute.org/investor]