BARRON'S

Your Personal Data Is Being Used by Investors. Here's the Potential — and the Risks.

By Avi Salzman   Updated Nov. 30, 2018 8:39 p.m. ET

Illustration by Alexander Wells

Even before the hip teenagers realized that *nobody* goes to the mall anymore, about a dozen math and science Ph.Ds packed into a co-working office in downtown Manhattan already knew it. On their computers, the Ph.Ds started to see a decline in the number of "pings" from cellphones that they were tracking in those malls. And that drop-off became an important signal for the stock market well before some malls announced disappointing financial results.

In a similar way, web-scraping company Thinknum in June noticed a drop in job listings on a Tesla site the day before the company announced a restructuring. And long before it became clear that Adidas was stealing market share from Nike and Under Armour, consumers who filled out online surveys created for investment bank Cowen & Co. started to say they preferred Adidas.

For savvy investors, such information presented buying and selling

opportunities.

How Fund Managers and Retail Investors Can Use Alternative Data

That kind of edge can be worth paying for, particularly at a time when traditional value-investing strategies are floundering and companies are withholding significant data from investors. Just a few weeks ago, Apple said it would no longer tell investors how many iPhones it sells each quarter.

The explosive growth of digital data, along with better tools to analyze and store it, has jump-started a once-niche industry that packages under-the-radar information for investors. Once standardized, the numbers can be plugged into computers, creating "big data" that helps investors predict.

While alternative data—like geolocation data that track the location of cellphones, or language-tracking software that detects changes in sentiment around a company—have been available for years, such data sets are increasingly becoming a more mainstream tool. Portfolio managers who have historically done most of their research using quarterly financial releases are now examining real-time feeds of credit-card transactions and social-media sentiment.

"This data is lying all over the place," says Michael Recce, who became the first chief data scientist at asset manager Neuberger Berman last year. "If we can analyze it and figure out who's really gaining market share, then we can make better investment decisions. Knowing exactly who is winning in the marketplace in real time is going to be a huge advantage."

Recce is convinced that with the right data sets, firms like his will be able to use information like credit-card receipts to "see Tim Cook's dashboard"—the company's most important statistics—even if the Apple CEO would rather keep it hidden.

Research has always been important for portfolio managers as an ancillary service. But now the stat guys aren't being put in a corner anymore. Neuberger expects Recce to eventually run his own portfolio.

Other companies known for fundamental investing processes are quickly staffing up, too. A new division at JPMorgan Chase (ticker: JPM) called Intelligent Digital Solutions is looking for coders and Ph.Ds who will work with asset-management teams managing $2 trillion. In December, it will hold its first "data science hackathon," bringing quantitative and fundamental analysts together with data scientists to crowdsource investment ideas using alternative data. It's a "paradigm shift" in how the firm invests, says Ravit

Mandell, the chief data scientist at J.P. Morgan Asset & Wealth Management.

The trend has even been embraced by some pension funds, historically among the most conservative investors. Marcel Prins, the chief operating officer of APG Asset Management, which manages the pensions of one in five Dutch families, says that using such data is now "part of being an active long-term responsible investor."

The push comes at a time when diving for data on consumers is under scrutiny. Facebook has been widely criticized for amassing and sharing enormous caches of personal information, stoking worries about data privacy around the world. So far, regulators have taken little action to stop investors from accessing alternative data, even when it comes from the credit cards and cellphones of consumers. But they may not stay silent for long.

"I find it hard to believe that eventually this intersection between privacy issues and insider-trading issues isn't going to be of some interest to folks like the New York attorney general or the Securities and Exchange Commission," says Jonathan Streeter, a former federal prosecutor who is now a lawyer at Dechert and one of the go-to legal experts in the field.

In his first few years in the private sector, he rarely heard the term alternative data from the asset managers he represented. About 18 months ago, he started receiving an influx of calls from fund managers wondering about the legal implications of using it. "The number of clients asking about it, and the number of events I'm asked to give speeches at, has ramped way up."

Most firms are just dipping their toes in the water. The average institutional investment firm spends about $900,000 annually on alternative data, according to a report from Greenwich Associates, which surveyed 40 investment managers around the world. Investors will spend about $300 million on these data sets this year, up from $170 million in 2017, the company projects.

That's a tiny sliver of the total value of financial data—spending on products from companies like Bloomberg and FactSet totaled $28.5 billion last year, according to Burton-Taylor International Consulting. But alternative data are undoubtedly making up a larger percentage every year. Bloomberg itself has been providing some of these data for almost a decade and has recently been adding new products, including satellite and geolocation information.

I find it hard to believe that eventually this intersection between privacy issues and insider-trading issues isn't going to be of some interest to

folks like the New York attorney general or the Securities and Exchange Commission. ”

—Jonathan Streeter, a former federal prosecutor who is now a lawyer at Dechert

The growing interest is also evident in the guest list for Discovery Intrepid, an annual conference put on by alternative-data company BattleFin on a decommissioned aircraft carrier in New York. Whereas, two years ago, 90% of the passes to the conference were bought by quants—hedge fund managers who use systematic computer-driven strategies—it's now "closer to 50/50" quants and more traditional asset managers, says Tim Harrington, BattleFin's CEO. The number of data providers presenting at the conference quintupled, to 153.

Overall, there are now more than 1,000 data providers charging anywhere from a few thousand dollars a year to several hundred thousand, experts say.

"The quants were very much the early adopters, and now there's a transition where the discretionary and fundamental world is waking up to it," Harrington says.

Harrington himself came from the world of traditional investing. He was an analyst at Steven Cohen's SAC Capital for years, covering telecommunications. He remembers driving to retail phone stores to talk to managers and see what was selling. Now, through BattleFin's platform, he sells a product that tracks every time a new phone is turned on and can find out quickly "which wireless carriers are gaining subscribers and which are losing them," he says.

Using alternative data is nothing new. For years, investors have used methods like channel checks—calls to a company's supply-chain partners—to gain an information edge. Even longer ago, Babylonian commodity traders would measure the depth of the Euphrates river to see which crops were likely to grow best, according to Ashby Monk, the executive director of Stanford University's Global Projects Center.

"Ultimately, alternative data is just data," says Monk, who has written extensively on the use of new kinds of investment data. "It's only defined as alternative because we actually took the time to define conventional. We define it as the thing that is not commonly used in investment decision-making but which could have value."

The new era of alternative data started after the financial crisis, says Leigh Drogen, the CEO of Estimize, which sells an alternate set of earnings estimates using its own survey. After 2008, the next generation of hedge

funds "were forced to go out and actually generate idiosyncratic alpha," he says. "Well, there ain't that much idiosyncratic alpha out there for people picking stocks." One idea they latched on to was to "try and leverage all this unique data," he says.

And the world has become awash in data.

"As a society, we produced more data last year than we did in the whole history of humanity," notes Tobias Moskowitz, a finance professor at Yale School of Management who is also a principal at Cliff Asness's quantitative investment firm AQR Capital Management. "That data is being captured all the time. Everything is recorded."

Gathering data today doesn't involve driving to every cellphone store in a state to ask how business is doing. Mundane activities—surfing the web, buying something with a credit card—now leave terabytes' worth of digital clues. Much of it is considered "waste" or "exhaust"—information that's created in the normal course of doing another kind of business. The people who make the weather or map apps on your phone track your location so they can give you accurate information about the weather. It just so happens that they can also sell that information to hedge funds, which use it to determine where people are shopping. There are hundreds of such apps that track locations with the permission of the people who download them.

Other companies then turn that raw data into useful forms. New York–based Thasos Group gathers anonymized location data from about 500 million phones that are running any one of more than 1,000 apps. The information comes into the company's computers as dots on a map, as in the case of the shopping malls. The firm's software links the signal counts associated with those mapped malls to the tickers of retailers or publicly traded mall owners, to see where traffic is growing or shrinking over time.

Thasos allows investors to see which mall customers are from wealthier census blocks, and which shop at multiple malls, among other insights. Bond investors can search by individual mall properties to determine if they're likely to stay creditworthy.

It's an incredibly laborious process, one that took Thasos six years to nail down. "It's like understanding particle behavior and trying to build a model," says Thasos CEO Greg Skibiski. "It's physics basically that we do here."

The company sells the full package of the product directly to investors for $100,000 to $200,000, and Bloomberg is offering a version of it directly through its terminal.

In July 2017, Thasos issued a press release showing foot-traffic data for the five top-performing and five bottom-performing real estate investment trust among the 30 largest ones that the company was tracking. The company's predictions weren't perfect—it said Simon Property Group (SPG) was lagging behind, but that company ended up increasing its earnings guidance—but the malls it said were winners all had positive price moves after reporting earnings, and three of the five companies it said would lag behind fell after reporting earnings. Several mall operators are themselves now buying data from Thasos, the company says.

Thasos has dozens of clients, but it's still relatively rare for most fund managers to pay for cellphone-location information, says Richard Johnson, an analyst with Greenwich Associates. About 10% of the money managers Greenwich surveyed used it.

"The location data gets a lot of headlines, but it doesn't seem to get a lot of usage," he says. That's not necessarily a bad thing, he adds. "The ones people aren't using now may have the most value," he says.

Geolocation information from another company called Alpha Hat has been helpful for Chuck Grom, a retail analyst at Gordon Haskett, a boutique research advisory firm. He used the information to write a note this year explaining to clients why Kohl's had made a smart move by partnering with Amazon.com to allow people to return Amazon goods to certain Kohl's stores. Stores where people could make Amazon returns had 8.5% higher traffic than stores where it wasn't available, Grom wrote in a note to clients. By using data on how much time each customer spent in the store, he projected that customers weren't just stopping by to make a return—they were actively shopping, too.

"It's a lot more of a refined approach than an image of cars in a parking lot," he says.

Among the most popular kinds of alternative data are those from credit and debit cards. Earnest Research compiles anonymized reports of spending by millions of consumers at hundreds of companies. It doesn't get every credit-card transaction, but a broad enough swath that "we have data on any company in the U.S. that accepts electronic payments," CEO Kevin Carson tells *Barron's*. The company keeps the exact source a secret. "We can't say specifically where we get it from," he says.

Earnest can track "essentially what you would see on your debit- or credit-card statement," he says, so it would include the size of the purchase and name of the company, but not exactly what the person bought. And it's not

just for retailers. Earnest can see how much people are spending on things like cable bills, too, Carson says.

Other providers gather spending information by interviewing people. Cowen & Co. has surveyed consumers for years to provide exclusive sentiment information for the research it sells to investors. Now, through a new entity called Kyber Data Science, it sells the information directly to investors.

Eyes in the sky are also watching. Orbital Insight, based in Palo Alto, Calif., uses satellites, drones, balloons, and unmanned aerial vehicles to snap photos, capturing information that would be tough to map from ground level. It's not just useful for checking how many cars are in the parking lot at Home Depot . Orbital says it can help answer crucial macroeconomic questions, like whether China is building more roads today than it did last quarter. Orbital also became available on Bloomberg this year and on CME DataMine.

Fund managers report mixed success at using satellite information. The investment firm AQR, for instance, has tested satellite data and found the benefits from them to be "fleeting" once the data became used more widely, Moskowitz says. "The data that gets a lot of hype is generally very fleeting. Markets figure it out pretty quickly."

Drones aren't necessary to gather some data. Web-scraping companies can derive information by building software that tracks changes to sites that post information like job listings or company sentiment information on sites like Glassdoor. That's how the web-scraping company Thinknum recognized that Tesla job openings had suddenly dropped in June, just before a restructuring was announced. Thinknum provides data on 400,000 companies, both public and private.

A similar company, YipitData, also collects and analyzes web information, but has narrowed its focus to 60 companies. By counting restaurants that partner with Grubhub and UberEats, it has compiled detailed data on who's winning the food-delivery battle, for instance. "If you're investing in Grubhub, you need that answer," CEO Vinicius Vacanti says.

Quantitative investors who plug all of this information into computers say they often backtest the sets for a year or more. At Two Sigma, another hedge fund known for strong results, the strategy can involve thousands of data sets.

"To be able to predict, you need to start with a ground-level understanding of the real world as it stands today," says Ali-Milan Nekmouche, chief data strategist at Two Sigma. "We use our platform to measure real world

economic activity in a granular way." Both Two Sigma and AQR caution that investors shouldn't look at these sets in isolation. Before inputting data it's important for a firm to have a "solid interpretation engine," Nekmouche says.

AQR starts with economic theories and plugs the data in to test its hypotheses and guide trading ideas. "I think it's very dangerous if you just take the data and throw it against the wall and see what happens," Moskowitz says.

Even the best predictive data can be tough to trade around. Web-scraped information that showed Tesla was slowing its hiring didn't predict how to profit off the information. An investor who shorted Tesla stock expecting a slowdown to worry Wall Street would have lost money: The stock actually rose on the day of the restructuring news.

Among the biggest wild cards in this industry is the threat of regulation. The SEC has cracked down on tools that give investors an edge, including so-called expert networks that claimed to offer insights from consultants. So far, alternative data have mostly steered clear of such scrutiny. Streeter, the Dechert attorney, says the SEC has been watching alternative data, though there have been no enforcement actions.

The SEC declined to comment for this article, and the New York Attorney General's Office did not respond to a request for comment.

Katya Chupryna, the chief strategy officer of Thinknum, says the industry has been serious about self-policing, banding together to create its own regulatory groups and come up with definitions of "personally identifiable information." She adds, "Because there are no standards, we are trying to build those standards ourselves, but always on the cautious side."

Hedge funds are wary of attracting regulatory attention and tend to vet data sets extensively, several industry insiders say. Unlike advertisers, they have no interest in targeting individual consumers for marketing purposes.

Streeter sees evidence that the industry is serious about self-policing, but he has looked at some situations that fell into a gray area.

"I have encountered situations where a vendor is offering to not only tell you how many cellphones were inside Best Buy on a particular day but also to tell you where the people live," he says. "And the way they figure that out is they follow those cellphones to where they go to bed at night and sit for 10 hours and wake up in the morning. That to me is a closer call, and I think that the law around that is undeveloped."

Even if such collection doesn't violate a specific legal principle, it "starts to raise more significant privacy concerns," he says.

Phone-location information has drawn some interest from legislators. Verizon, AT&T , Sprint, and T-Mobile U.S . all limited the location information they provide to data brokers in June after criticism from Sen. Ron Wyden, a Democrat from Oregon. That didn't stop phone companies from disseminating some information to third parties, and it didn't stop apps from tracking phones.

Wyden wants to go further. He released a draft of a bill in November that would give consumers more power over the personal information that corporations collect, including location data. "There is clearly a vast, largely unregulated market where companies, data brokers, asset managers, and others swap and sell Americans' personal information," he said. "This market is sorely in need of radical transparency, real oversight, and tough penalties for those who misuse our data or lie about protecting our information."

Thasos can pinpoint the census block area where a cellphone's signal originates, but it makes sure that all personally identifiable information is scrubbed from the information before it enters the company's system, CEO Skibiski says. That condition is in its contracts with all of the apps it works with, he says. "I'm super-privacy-sensitive myself," he says.

Other data sets have also generated some scrutiny. In 2015, Massachusetts Institute of Technology researchers said they were able to identify people using a small amount of information in credit-card purchase databases, even when those purchases were made anonymous. Kevin Carson, of Earnest, says that its products "are delivered to clients in aggregate form, and do not contain any personal information."

Privacy advocates haven't been particularly focused on alternative investment data, but see echoes of other recent controversies in the field.

"Hedge funds are seen as bad actors in other ways, but are not seen as the biggest baddest privacy actors yet," says Gennie Gebhart, associate director of research at the Electronic Frontier Foundation. Making data anonymous is easier said than done, she says.

"It's incredibly hard to make a truly anonymized data set" at a time when there's a growing business in deanonymizing consumer data, she says. "Anytime you're combining disparate data sets, there is the risk of creating new privacy-invasive information or assumptions."

Write to Avi Salzman at avi.salzman@barrons.com

A Better Approach to ESG?

Investing for Profits and Social Benefits

Just How Gender-Equity Focused Is That Fund, Anyway?