ALPHAVILLE

HOME MARKETS LIVE LONG ROOM

Equities Hedge funds Markets

In defence of hedge-fund data mining

SEPTEMBER 13 2016 3:52 PM
By: Alexandra Scaggs

People tend to get worked up about the idea of Wall Street mining the trails of data (http://www.wsj.com/articles/wall-streets-insatiable-lust-data-data-data-1473719535) we all leave on the internet for investment ideas.

The first and most obvious reason is that privacy issues are always contentious. You'd think it'd be a windfall for hedge funds if MasterCard or Visa started selling them data on trends in customer purchases (while the card providers sell some data (http://www.mastercardadvisors.com/information-services.html), hedge funds haven't cracked that yet, as far as we know). While the information obtained in data-mining isn't supposed to be traceable to individuals, how would a company fully scrub that of any signs of consumers' identities?

A second reason is the sheer oddness of the regulatory environment that makes so-called "alternative data" OK to use, as Matt Levine points out (https://www.bloomberg.com/view/articles/2016-09-13/fake-accounts-and-artisanal-data).

The information is proprietary — as in, a mom-and-pop investor couldn't get the same thing — but it's fine to trade on (http://www.casebriefs.com/blog/law/corporations/corporations-keyed-to-hamilton/transactions-in-shares-rule-10b-5-insider-trading-and-securities-fraud/chiarella-v-united-states/2/) because it hasn't come from the company, or from someone with a special confidential relationship to the company, like a banker on a deal. (Meanwhile, the Supreme Court is scheduled to hear a case (http://www.scotusblog.com/case-files/cases/salman-v-united-states/) about whether trading on information from insiders who don't get paid for their tips still constitutes insider trading. What a world.)

But there's also one reason we *shouldn't* worry too much about alternative data: it doesn't scale (https://www.mcsweeneys.net/articles/when-it-comes-to-mazes-for-people-to-wand er-endlessly-before-starving-to-death-we-cant-be-beat).

That's why venture-funded shops like Quandl, which just closed a $12 million Series B financing round, are so interesting. They don't quite fit with the "information-wants-to-be-free" ethos of Silicon Valley.

The entire point of alternative data is to give hedge funds an edge over competitors, which analyse increasingly expensive (https://www.ft.com/content/785092ec-33d8-11e6-ad39-3fee5ffe5b5b) public data, or even use interrogation tactics learned from former spies (https://www.ft.com/content/a33eb562-0d61-11e6-ad80-67655613c2d6). If a hedge fund can extrapolate information about new car sales from car-insurance policies, they want to do so *before* Ford or GM or Toyota releases sales numbers, when the rest of the industry trades on them.

The edge given by exclusivity will evaporate the moment any good source of alternative data is widely disseminated. The early data adopter then becomes the investing world's equivalent of a frustrated hipster whose favourite band just sold out Madison Square Garden. So hedge funds jealously guard their proprietary data — and we suspect they have the cash to pay up for top-notch security.

Quandl's Kamel said the company is careful when deciding how broadly to disseminate its data sets. Instead of looking for more funds to buy its data, it's looking for new sources of "data exhaust (http://www.wsj.com/articles/wall-streets-insatiable-lust-data-data-data-1473719535)," or unused data created by a business's operations, and new methods of analysing that data. (That's where the artisanal methods of collection (https://www.bloomberg.com/view/articles/2016-09-13/fake-accounts-and-artisanal-data) discussed by Matt Levine come in.)

"We believe that if the data's public, any alpha that's there has been tapped long ago," said Kamel. "We want to preserve the alpha for as long as possible. But I'm not naive enough to think dissemination won't happen in time, which is why we need to keep our pipeline always full. So we have to keep finding new and powerful data for our customers."

Tripp Jones, a general partner at August Capital, which led Quandl's first round of fundraising, acknowledges it's a tough line to walk.

"One, I think alternative data is so broad, by definition, that you have to know what you're looking for. Two, it can be solved with price. The more valuable the data, the more expensive it will be," said Jones. Even so, "I think a lot of alternative data will be next to free."

That is a less-than-comforting response from a privacy perspective.

But, well, we've all basically given up on the idea of digital privacy, right? From that (only half-serious) perspective, it would seem that the ways our data have been used — targeted advertising (https://techcrunch.com/2016/08/25/whatsapp-to-share-user-data-with-facebook-for-ad-targeting-heres-how-to-opt-out/), government surveillance (https://www.theguardian.com/us-news/the-nsa-files), credit monitoring — are more invasive than the goals of the hedge funds doing the mining.

Think about the end goal of each type of surveillance: Advertisers want your money. The government wants to figure out if you're a security risk, or if you know any people who are. Lenders want to know if you deserve their money. The hedge funds just want to figure out *where* you're spending your money, so they can put their money in those places too.

So, what's the problem if a hedge fund sees a rise in pet-care spending, and buys PetSmart shares in response? Or, if they notice people are abandoning a retailer en masse, why shouldn't they bet on a share-price decline? That's an allocation of capital to companies where people actually spend their money, which is better than (https://www.bloomberg.com/view/articles/2016-08-24/are-index-funds-communist)~~socialism~~ indexing.

In fact, why not just set up a system where people can sell their information to hedge funds themselves? If average people are going to be watched regardless — again, we're not entirely serious, but we're also not entirely wrong — why not compensate them for it?

Related links:

Wall Street's Insatiable Lust: Data, Data, Data (http://www.wsj.com/articles/wall-streets-insatiable-lust-data-data-data-1473719535) – WSJ
Fake Accounts and Artisanal Data (https://www.bloomberg.com/view/articles/2016-09-13/fake-accounts-and-artisanal-data) – Matt Levine
When it comes to mazes for people to wander endlessly before starving to death, we can't be beat. (https://www.mcsweeneys.net/articles/when-it-comes-to-mazes-for-people-to-wander-endlessly-before-starving-to-death-we-cant-be-beat) – McSweeney's Internet Tendency
Twitter Paid $134 Million for Data Partner Gnip (http://http://blogs.wsj.com/digits/2014/08/11/twitter-paid-134-million-for-data-partner-gnip/) – WSJ
The Billion Prices Project: Using online data for measurement and research (http://voxeu.org/article/billion-prices-project) – Voxeu.org