

The Promise and Peril of Big Data and Machine Learning on Wall Street

Abstract:

Because the methods and processes for investing with hedge funds is deregulated, a number of high tech firms have turned to machine learning to help them gain greater insight into potential investments. Through the mining of unconventional data from a variety of sources, hedge funds have been better able to use sentiment analysis, ESG data, geospatial data, fundamental, and technical analysis to provide higher return to their shareholders. Caution, however, should be taken before assuming that these methods can sustainably provide higher returns that traditionally managed hedge funds and market benchmarks.

INTRODUCTION: Here's how the story typically goes: A young ivy-league graduate goes to work as an analyst in a bulge bracket investment bank such as Merrill Lynch or Goldman Sachs. After putting in their two years, getting yelled at for having their excel sheets right-aligned instead of left-aligned, and being on call 24/7 to be able to whip up a merger report at 2AM after a weekend night out in a dive bar in New York City, the most promising of these young analysts will move on to more prestigious roles in venture capital, private equity, or in some cases, the mysterious world of hedge funds. The first hedge fund was created in 1949 by Alfred Winslow Jones, but it was only since the 1990's did they begin to proliferate. Hedge funds are vast, unregulated pools of money from high net-worth individuals and institutional investors seeking a return much greater than what they can find from a mutual fund. Because hedge funds are, for the most part, unregulated, they can invest in a wider array of options and utilize riskier strategies such as M&A arbitrage, options, distressed debt, and emerging market equities. Furthermore, unlike mutual funds, there are no regulations dictating a set level of diversification in a hedge fund. While this lack of regulation allows some hedge funds to deliver a large return, it also leaves investors without protection should the fund go south. While both individual and institutional investors

have protection with the FDIC and SEC with mutual funds, these funds are also restricted from owning certain equities, taking excessive risk, or employing particular strategies. Investors began to see the inefficiencies in the market that resulted from these arrangements and stepped in, creating limited partnerships with less than 99 investors, just below the SEC threshold to be regulated. Today there are around 16,000 hedge funds in existence.

While the most talented fund managers are able to make a considerable return, too many are not. The HFRI Fund Weighted Composite Index has shown just a 1.7% annualized return over the last 5 years, and even that number is likely to be inflated, considering that the hedge fund industry suffers from return manipulation and survivorship bias. This low performance, combined with the exorbitant fees that come with management, has steered many asset managers away. In the first part of 2016, \$15.1 billion dollars left hedge fund markets, much of that going to passive investments such as index funds. Just 2 years ago, California's largest pension fund bailed out of hedge funds, citing shrinking profitability and excessive cost. It is hard to blame these managers when passive investment vehicles have been outperforming actively managed large cap equity funds. It is obvious that if these types of funds want to stay afloat, they need a new edge to be profitable.

According to a paper by Cap, Goldie, Liang, and Petrsek, one of the reasons that fund managers are so well compensated, with high performance based and management fees, is that talented fund managers are better able to manage downside risk than other institutional investors. The authors found that "the source of hedge fund outperformance is not a hedge fund's ability to identify the best deals in which to invest, but rather its ability to avoid the worst deals." This is in line with the role of hedge funds, that is, providing superior returns at similar levels of risk, and hedging against a stagnating economy. It is easy to see then, how a machine learning strategy could be useful to fund managers. By utilizing an algorithm that is able to make a constantly updating model of losing strategies and the ingredients for success, these algorithms will be able to detect micro patterns from large swaths of past

and present data. Managers who use these strategies, will be much more equipped to avoid the volatile downswings of the market.

Since early 2000s, with the breakthroughs in artificial intelligence and machine learning, cutting-edge hedge funds have been hiring an increasing number of engineers, physicist, computer scientist, and mathematicians to create their investing strategies. While many of these strategies use bread and butter fixed algorithmic models to make automatic decisions based on past market trends, a growing number of hedge funds have utilized machine learning to make models that constantly update themselves with new information and sources of data.

While the popular perception of algorithm-wielding money managers is the 25-year-old college dropout utilizing High-Frequency Trading to take advantage of the short-term price spreads created from the volatility of the stock market, this image has shifted in recent years with the introduction of nascent machine learning based strategies that can predict stocks that have long term value. Rebellion Research is able to use data to make decisions about what stocks will generate the most upside over a period of two to three years. The theory is that Wall Street and investors are quite terrible at learning from their mistakes, while machine learning model is quite adept at constantly refining its strategies based on the correctness of predictions and new market data. In the case of RR, their strategy has worked. Since its launch in 2007, they have been able to top the S&P 500 by almost 40-50% in cumulative returns over the last 9 years (after fees). Rebellion Research is not the only hedge fund jumping on the band wagon. Over the past few years the number of quant hedge funds that sponsor the Conference of Neural Information Processing Systems – a machine learning and computational neuroscience conference – has quickly grown. 15 out of the 62 sponsors work in the financial services industry, and there seems to be strong evidence that the number of firms utilizing neural networks and machine learning will only increase. Alexander Fleiss, the founder of Rebellion Research, believes that machine learning is the future of investing, and has no reservations about it.

REBELLION RESEARCH: Created by Alexander Fleiss and three of his friends in 2005, Rebellion Research is an asset management firm (read: mutual fund) devoted to bringing machine learning and artificial intelligence based strategies to investors. Their “proprietary system...predicts future returns while weighing potential volatility.” According to their pitch book, their A.I. draws data from hand-selected macro, fundamental, and technical factors to get a good handle on a company’s growth, value, and momentum. The focus at RR is not short-term, high frequency algorithmic trades based on cyclical short term patterns in the market. Instead, they use Bayesian statistics to match prior market and performance knowledge with new data to predict stock performance. These models will allow the investors to see the factors that will most effect stock performance and create a probability distribution of potential returns. Their strategy, for the most part, works. From their genesis in 2005, they have grown to almost 20 billion assets under management, and have a cumulative return of almost 125%, compared to Credit Suisse’s AllHedge benchmark of around 7% and S&P 500 cumulative return of around 75% during the same time period. Much of this return can be attributed to their data driven approach. Monitoring daily economic data from over 53 countries, monitoring the strength of GICS sectors (ten of the most influential sectors in the market) and looking at relevant factors including recent M&As, customer demand, commodity prices, and changing COGS, Rebellion Research can pick out the strongest positioned companies from a fundamental standpoint, looking at the balance sheet, earnings, and cash flow. Rebellion Research then uses its proprietary A.I. to find the optimal holding size given expected risk and return.

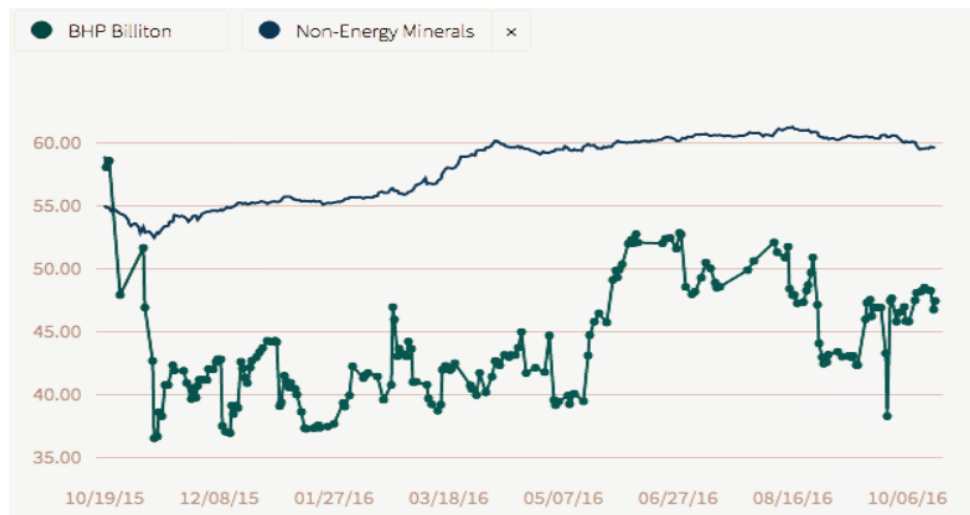
Beyond choosing the best assets to include in their portfolio, they also use their A.I. to properly hedge risk. The model was built to limit volatility and reduce the *beta* of any one asset class (reducing the deviation of holdings in one asset class from total asset class return, aka diversification). In addition,

the A.I. is also used to limit credit, political, currency, and interest rate risk. The combination of these strategies employed by RR to reduce risk and volatility means a smoother return curve and less down side.

TWO SIGMA: Two Sigma, another hedge fund that has been utilizing more STEM majors than Finance majors in building their investment strategy, has \$37 billion AUM and plenty of brain power at their disposal. Founded in 2001, they operate foremost a technology firm, and secondly as a finance investment house. 2/3 of their staff is research and development, and 60% of their hires are from non-finance backgrounds. Their models utilize 10,000 unique sources of data and their investment models perform trillions of calculations per second. From 2005 - 2014, their Compass Enhanced fund has returned over 30%. One of the most preeminent quant players in the hedge fund space, they continue to attract top finance and academia talent to build their impressive models.

ESG DATA: While many funds are using past market performance and prices to make models to make forecasts into the future, there is a different camp that uses non-financial Big Data to make more responsible and less risky investments. According to a recent article by the *Financial Times*, more than 40% of hedge funds are warming to the idea of using ESG data for responsible investing. Several funds even believe that ESG factors - economic, social, and governance - have a measurable financial impact. Through machine learning and mining usually inaccessible data, they have been able to find meaningful correlations between social and governmental responsibility that measures and construct portfolios that generate high returns over the long run. Some of the factors that they measure include climate change, emissions, human capital, community relationships, corporate structure, and executive compensation. While historically, most of these factors were passed over for concrete business metrics, investors are increasingly seeing the advantage of using sustainability data. The belief is that underperformance on

ESG metrics could have a detrimental effect on earnings and share price over the long run. Insight360 by TruValue Labs has a constantly updating ESG data platform that tracks a corporation's environmental, social, and corporate governance record for both individual companies and sectors overall. In a case study posted on their blog, Insight360 was able to show that BHP Billiton's ESG score (out of 100) steadily declined throughout 2015. This mismanagement crescendoed into a nightmare on November 5, when a tailings dam collapsed and killed 19 people and turned the river into a toxic mudslide. BHP Billiton's stock price dipped to an all-time low of \$20.19 just two months after the incident.



SENTIMENT ANALYSIS: Creation of social media sites and news wires have provided computers the data set they need to automatically mine information from both the market and the population at large. Sentiment Analysis, the extraction of subjective information from source materials, can be used on news headlines, twitter data sets, or discussion boards to glean relevant information, as well as general sentiment about the market. Through these techniques, traders can know beforehand, when their investments, or potential investments, are likely to drop in price due to bad news about earnings, scandals, macroeconomic trends, or regulation. While rudimentary forms of sentiment analysis have

been around for a while, the field has recently been given a boost with advances in natural language processing and deep learning.

GEOSPATIAL DATA: Finding patterns in historical stock prices among Big Data is not the only way that hedge funds are leveraging their newfound computing abilities. Startups, such as Planet Labs and Orbital Insight, give hedge funds and other investor types the geographical and pictorial data, along with the machine derived analysis of this data to make informed investing decisions. Thanks to a fleet of shoebox sized satellites that orbit Earth in a constant pattern, Planet Labs is able to sell global imagery to hedge funds, NGOs, governments, farmers, and businesses to lever the advantage of new insights. Planet Labs also sells their data to a company called Orbital Insight that integrates Planet Lab's and other satellite provider's geospatial data with machine learning algorithms and parallel computing to sell "actionable intelligence for the world" Orbital Insight's reports have been used from everything from quantifying a China Economic Index through calculating factors such as car density, building construction, port activity, commodity stockpiles, agricultural yields, oil storage, and coal plant development and using surface area of water bodies to estimate global water reserves. The combination of new satellite technology and data mining processes have come together to provide insights that were previously too cumbersome and expensive to find.

POTENTIAL CONFLICTS: While a recent KMPG study shows that 58% of fund managers say that machine learning will have a medium to high impact on the way hedge funds operate, there is a reason to believe that these methods might not be as successful as it might seem. Investing in the algorithms and investment strategy up front requires an enormous amount of upfront time and capital to get the project off the ground. It will take more than just a whim of a former banker, a few friends from Harvard, a few million dollars to invest. Furthermore, the failure rate of new algorithms in live trading

situations is close to 90%. Millions can be sunk into hiring mathematicians and software engineers, only to produce an unprofitable algorithm that only works in theory. According to David Siegel, the founder of Two Sigma, “machine learning systems can easily, with high confidence, make mistakes.”

Unfortunately, the only way to see if you have a winning algorithm or not is to actually make one and see how it performs in a real live market. This, as we know, puts millions of dollars of investor money on the line, not to mention the capital that it took to get the system up and running.

Secondly, the major advantage that a machine learning based system has over a human trader is its ability “to look for patterns across hundreds of markets”. Compared to a human trader that can only manage a dozen or so positions, a computer has the computational ability to pull data from hundreds of sources at once, monitor the positions a firm is currently in, and update its model simultaneously for each position as new information comes in. Beside this numbers advantage, hedge fund manager Douglas Greenig says “a good human trader will still usually outperform an algorithm in a single market.”

Like computer algorithms that learn to discriminate against people-of-color when reviewing applications for home mortgages, trading algorithms can fixate on irrelevant variables and correlations to make predictions as well. For example, if a machine learning algorithm is given too much data to process, it may find a meaningless pattern between say US labor productivity and the S&P 500. While this relationship seems intelligent, there are a lot more factors that affect the S&P 500 index, and it could be a mere coincidence that these two numbers rise and fall with each other. If you were to adjust the time frame in which the algorithm searched for patterns, you are likely to find that there are instances in which these variables aren't as related as they seem. In 1997, money manager David Leinweber tried to find the golden variable that best predicted stock performance from 1981 to 1993. After running through thousands of variables, he finally found the best fit: the total volume of butter produced each year in Bangladesh. While we can be confident that a smart A.I. algorithm trying to find

patterns in seemingly meaningless data would find quickly that butter output is NOT a good measurement of performance, it still might make a decision in a more plausible, yet similarly fallacious way.

Finally, there is reason to believe that machine learning driven funds might be enjoying success in the present because they have done what the first hedge funds were able to do; exploit inefficiencies. Once more and more firms start moving into the same space, the competitive advantage that firms such as Rebellion Research and Two Sigma enjoy today could be dramatically reduced or destroyed. It is not an implausible idea that several similar algorithms finding the best, undervalued investments, will quickly dry up all of the lucrative investment opportunities, and all assets will be priced at fair value, much like the Efficient Market Hypothesis suggests.

CONCLUSION: It seems that the avalanche of data is able to help Wall Street in numerous ways, but it is still too early to tell whether this data will be able to sustainably drive greater returns than traditionally managed funds. Even the managers of the quantitative and A.I. based funds don't discount the role that human talent plays in the management of successful funds. It seems like the in the world of finance, intuition still has a place on the stage.

REFERENCES:

Jason Sweig, 2007. *Your Money and Your Brain How the New Science of Neuroeconomics can Help Make you Rich*. SIMON & SCHUSTER.

Nishant Kumar and Taylor Hall. *Why Machines Still Can't Learn So Good*. Bloomberg.

Charles Cao, Bradley A. Goldie, Bing Liang, Lubomir Petrusek 2016. *What is the Nature of Hedge Fund Manager Skills? Evidence from the Risk-Arbitrage Strategy*. *Journal of Financial and Quantitative Analysis*. Vol. 51, No. 3.

"Home." *Planet*. N.p., 14 Nov. 2016. Web. 17 Nov. 2016.

Rebellion Research, 2016. *Rebellion Research 2016 Presentation*.

Sigma, Two. "Discovering Value in the World's Data." *Two Sigma*. N.p., n.d. Web. 17 Nov. 2016.

Matt Egan, "The part is over for hedge funds and the hangover could hurt" May 18, 2016, CNN Money Foundation, NIPS. "2016 NIPS Sponsors." 2016 NIPS Sponsors. N.p., n.d. Web. 17 Nov. 2016.

Caroom, Eliot. "BHP Billiton Dam Disaster Began a Terrible Year of ESG Performance for Mining Giant." *TruValue Labs*. N.p., 19 Oct. 2016. Web. 17 Nov. 2016.

Reynolds, Fiona. "Subscribe to Read." *Subscribe to Read*. *Financial Times*, 16 June 2015. Web. 17 Nov. 2016.