Machine Learning Meets Portfolio Management

Shane Obata, Chris Kerlow, Craig Basinger, Benjamin Ye

> *This Market Ethos is an excerpt from two larger reports published by the Connected Wealth team. The first is an introductory primer on Big Data and Machine Learning (full report <u>HERE</u>). The 2nd is a White Paper on how we used machine learning to better optimize an investment strategy that targets investor herd behaviour around analyst recommendations (full report <u>HERE</u>).*

Financial markets have become increasingly efficient, making it harder to beat them using conventional methods. While we are still in the early days, some professionals are already using the wealth of available data to apply sophisticated quantitative techniques such as machine learning. These techniques will never solve the market; but, they may help money managers to generate alpha in an increasingly competitive industry.

Big Data and the Future of Finance

We are in the midst of an exponential rise in data. IBM estimates that 90% of the world's data was created in the past two years. The flood of data is coming from three primary sources: individuals, companies and sensors. Individuals are contributing with every Instagram post & Uber hail, companies are generating more transaction data than ever before, and people are putting sensors into everything from washing machines to wind turbines. Global data generation amounted to ~21 ZettaBytes (ZB) in 2017 and is expected to rise to more than 160 ZB by 2025, according to Seagate.

The financial world is also awash in data. Financial markets, economies and news provide us with a near limitless supply. Most investors have access to the same data at around the same time. This is making it harder to beat the market using conventional methods. In response, some money managers are exploring alternative data sources as they try to find an edge. For example, some investors are examining satellite imagery to assess oil rig & shipping activity and credit card transactions to evaluate sales trends. Others are analyzing social media to gauge sentiment. The 2nd chart on the right is a Bloomberg Twitter Activity Chart of Facebook (FB) from December 29th, 2017 to April 3rd, 2018. The top panel shows the Twitter publication count and the price while the bottom panel shows the sentiment breakdown. As we can see, there was a dramatic increase in the publication count, and more specifically negative sentiment, following the Facebook-Cambridge Analytica data scandal, which involved the collection of personal data from up to 87 million FB users. The big data revolution has been driven by advancements in computing power and storage and by falling costs top chart on page 2. PCs are faster than ever before and our iPhones store as much

Global Data Generation (in ZB)

Note: 1 zettabyte (ZB) = 1 trillion gigabytes (GB)

Source: Seagate
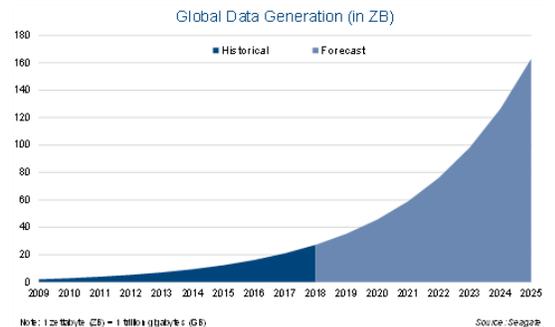
as old computers did. Access to and use of data has also improved dramatically. Public, private and hybrid clouds are allowing multiple users to access data simultaneously, with limited overlap of storage. More usable data has brought about novel applications. Autonomous vehicles, facial recognition and the native ads in your Facebook feed are just some examples. Many industries are using data science to advance their businesses. Finance is one of them; however, it has been a relative laggard, particularly in Canada. The main reason is that there are not many professionals who possess both programming and finance skills. The applications of machine learning in finance is wide spanning and so are the types of algorithms you can use to implement these strategies. For those who want to learn more, we have published a report, click HERE, which is a deep into big data and machine learning.

Cost of Computing - Falling Since 1960

Notes: *GFLOPS = GigaFLOPS - 1 GFLOPS = 10^9 FLOPS. The graph has been logarithmized to a base of 10.   *Source: Wikipedia*
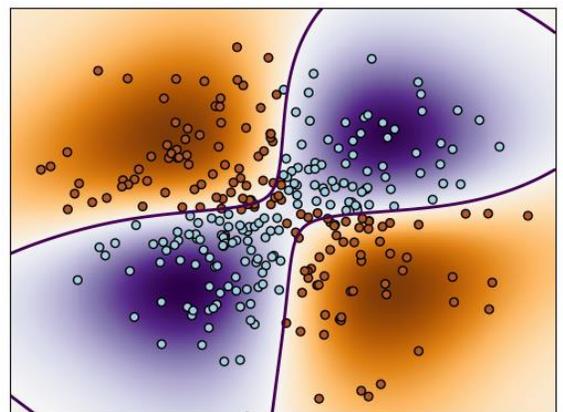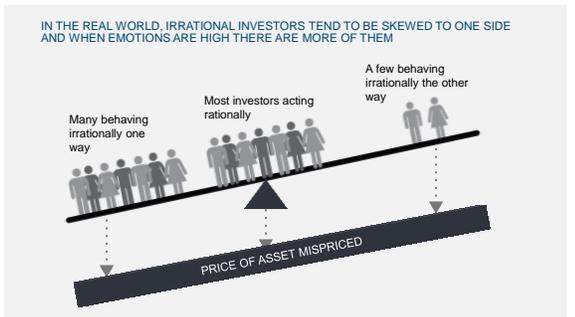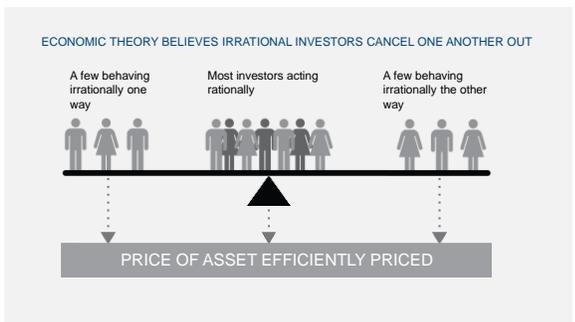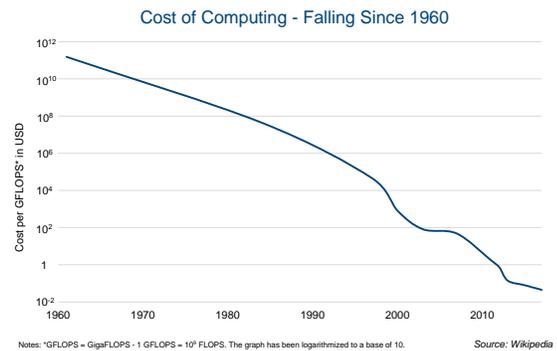
Applying Machine Learning (ML)

We are not true data scientists by training but recognize the opportunities this discipline has to offer. Thus, have hired some "quants" and begun to implement some techniques into our portfolio management regime to aid our process and hopefully enhance returns. Our thought is that human + machine is greater than either in isolation. We are also firm believers that behavioural biases – driven by emotions – cause investors to act irrationally. This can lead to mispriced assets. These biases are human in nature. As a result, they should continue to elicit bad behaviours that result in predictable mispricing. Using a combination of quantitative, emotionless strategies "machines" and our own experience in behavioural finance and traditional portfolio management (human aspect) we try to identify mispriced assets caused by investors acting irrationally in one direction or herding and profit from it.
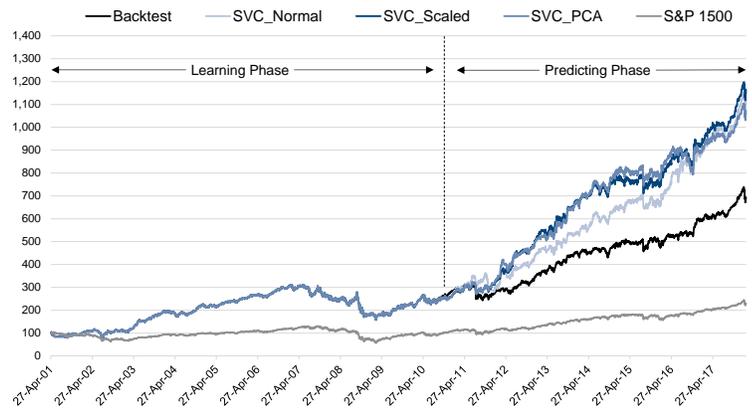
In our second report, found HERE, we share our analysis on how to profit from the herd behaviour elicited by analyst recommendations. We also demonstrate how applying machine learning can help to improve upon a base strategy.

ECONOMIC THEORY BELIEVES IRRATIONAL INVESTORS CANCEL ONE ANOTHER OUT

A few behaving irrationally one way

Most investors acting rationally

A few behaving irrationally the other way

PRICE OF ASSET EFFICIENTLY PRICED

An Example of Strategy Development with ML

The base approach for our Unloved to Less Unloved strategy was good before we started refining it; our goal was to make it better. The first step was to create a screening tool. Excel was too slow, so we partnered with Bloomberg's quant team to develop a system that performed most of the work in the cloud. The result was a dynamic Python-based tool that allowed us to screen various indices for trading instances. We used this tool to compile a database that goes back 20 years. The second step was to automate our trading rules, which involved setting ideal stop-loss and reset (effectively trailing stop) levels. Optimization greatly improved the results of the unloved strategy, which outperformed the S&P 1500 benchmark over the entire data range. The third step was to apply machine learning. We used a facet of machine learning called supervised learning where we provide the machine with both the question (X) and the answer (Y). It then learns from the data by identifying patterns, using whatever algorithm we choose to apply. We tried using various models but settled on non-linear Support Vector Machines (SVMs), since they consistently performed well. SVMs are a set of supervised learning models that tend to work well in high-dimensional spaces. These machines are used to separate data into classes using "decision boundaries." The right figure shows how a non-linear SVM classifies blue versus brown points with a high degree of accuracy based on such boundaries. For our purposes, the aim was to separate the trading instances into YES or NO classes, yes being trades that produces alpha (excess returns) no, trades that do not. The results were very promising, as shown in the chart on page 3.

IN THE REAL WORLD, IRRATIONAL INVESTORS TEND TO BE SKEWED TO ONE SIDE AND WHEN EMOTIONS ARE HIGH THERE ARE MORE OF THEM

Many behaving irrationally one way

Most investors acting rationally

A few behaving irrationally the other way

PRICE OF ASSET MISPRICED

Another interesting aspect of machine learning is that back tests are not just basic back tests, that are refined until you get the result that you want. In this example we used ten years of data to let our machines "learn". This part is like a simple back test where you refine and test to see what works best. Then we used the latest, roughly 8 years, of uncontaminated data to see if our predictive algorithms would actually add value. Running the machine on a daily basis, looking for trading signals, running those trades through the algo, and buying only those that are predicted to add value and not buying the rest. You can see that the predictive algos have done significantly better than our rudimentary model.

Machine learning is often framed as a black box. Inputs go in, magic happens, and outputs come out. That is not how it actually works. The models have improved over time and are now more transparent than ever. We determined which features (variables) were the most important by running a series of simple functions. Some of the features that came up repeatedly were fundamental revisions, trend scores (momentum) and lagged returns.

Backtested data is for illustrative purposes only. Please see the disclaimer on back testing below

Conclusion

The rise in availability and use of data are already having an impact in finance. It is still early. However, some professionals, us included, are already using the wealth of available data to apply sophisticated techniques. Machine learning will never solve the market. Even so, it may help practitioners to find an edge in an increasingly competitive industry. We are optimistic that data science will continue to add value in our investment process and that the behavioural mistakes made by humans, causing asset mispricing, will not go away. We think that combining traditional portfolio management, behavioural economics and machine learning techniques should help us gain an edge.