Bursting the big data bubble

In the financial world, big data is hailed as a potential game changer for predicting stock market performance. But without adequate safeguards, big data analyses may result in spurious correlations, misguided predictions and disappointing returns. By **Wai Mun Fong**

People cannot help looking for patterns, even in areas of life where randomness rules. Nowhere is the search for patterns more intense – and futile – than in the daily drama of stock market prediction. Game plans and strategies formulated by traders are predicated on the idea that it *is* possible to stay one step ahead of the stock market, that predictors *are* out there – one simply has to know where to look.

This belief persists despite years of research demonstrating that it is next to impossible to accurately predict stock returns for *any* data interval, be it minute, hour, day, week, month or year. The big picture is even more disheartening. As Stephen Ross, a leading finance researcher, points out, it is hard to even explain stock market movements after the fact using any set of observable fundamental variables one chooses to dredge from the available data. He writes: "It is one thing not to be able to predict what asset returns will be … it is another … to observe the movement of prices and not know why they moved after the fact."[1]

The failure to predict stock market movements is not for lack of trying. A vast number of observable fundamental proxies have been used in studies, including past returns and trading volume, equity issuing activity, dividend yield, interest rates and spreads, inflation rates, book-to-market ratio, stock market volatility, investment–capital ratio, and consumption–wealth ratio. Consistent with Ross's observation, a comprehensive study by Goyal and Welch[2] of equity prediction models found that many results were not statistically significant "in sample" – that is, the models were not able to make accurate predictions of the stock market data they were trained on. Not surprisingly, most of the same models performed poorly when applied to fresh "out-of-sample" data. Complementary to these results, Goyal and Welch show that a "kitchen sink" in-sample regression of annual returns for the S&P 500 index, using all the above variables, produced a coefficient of determination of just 13.8%, a telling indication that stock returns are too noisy to predict even at the annual frequency.

Other prominent studies[3,4] support these general findings – but traders, and researchers, are not to be deterred. And so it is that in recent years "big data" has been talked about as a potential game changer for stock market prediction.

Online footprints

Investors, like regular people, make frequent use of the internet: shopping and streaming and searching for whatever takes their interest. All of these activities generate data – and this data makes up a person's "digital exhaust" or "footprint". Much of this data is then made available as huge, aggregated data sets, and it is these big data sources that hold such hope and promise for those looking to predict stock market performance.

Wai Mun Fong graduated with a PhD in financial economics from the Manchester Business School in 1992. He is currently Associate Professor of Finance at the National University of Singapore Business School where he teaches personal finance and wealth management. His research focuses on topics in asset pricing, market anomalies, behavioural finance, portfolio analytics and personal finance

Main image: World Image/bigstock.com

> Years of research show that it is next to impossible to accurately predict stock returns – but big data has been talked about as a potential game changer

The idea underpinning the hype seems reasonable: the online activity of people who have an interest in the stock market may be related to their actual trading decisions, thus the "footprints" they leave behind may be useful for predicting where stock prices are heading. For example, when an investor sits down in front of a computer screen and opens their web browser, the terms they search for may reveal something about their current interests or concerns. If many investors sit down and search for many of the same terms, search volumes for specific keywords will spike. If search volumes increase for *debt* or *unemployment*, say, this may indicate that investors are nervous about the strength of the economy, which in turn may reveal something about their short-term investment strategies. From this, it may be possible to predict the movement of stock prices.

Google is an obvious place to look for these "tell-all" search terms. The company has the largest market share for keyword search volume (according to comScore, which measures website usage), and since 2004, Google has made its vast database of keyword searches available to the public, free of charge, through its Google Trends website. Google Trends compiles a search volume index to measure how often a particular keyword or phrase is searched, relative to the total search volume in various regions of the world. The search volume index is normalised to have values ranging from 0 (meaning the search volume does not meet a designated threshold) to 100 (the highest relative search volume at the time). The data is available weekly.

Google itself has exploited this vast trove of data to try to make real-world predictions. In a 2008 *Nature* paper, it claimed that by monitoring searches for flu-related terms, its Google Flu Trends detection system could predict the proportion of people in the USA having the flu bug in a more timely and accurate way than traditional epidemiological surveillance data, such as those compiled by the Centers for Disease Control and Prevention (CDC). A year later, Google had to tweak its models after they badly underestimated the outbreak of the H1N1 (swine flu) outbreak. Then in 2013, its models overestimated the peak of the flu season by a whopping 140% relative to CDC data (bit.ly/2oJb5HX).

The Flu Trends failure highlights how easy it is for wild-card factors, such as unexpected flu strains or changes in people's search behaviours, to confound algorithms. But perhaps the key lesson here, for those looking to relate search data to stock market performance, is the most obvious one: search activity may only be loosely correlated with real-world events, so predictions based on this may be doomed to fail. That lesson has yet to be learned.

Trading signals

In a *Scientific Reports* paper published in 2013, data scientists Tobias Preis, Helen Susannah Moat, and H. Eugene Stanley[5] used Google Trends data to test whether search volumes for a large set of finance-related keywords could predict the historical performance of the Dow Jones Industrial Average (DJIA). This index, based on the performance of 30 large companies, is widely used as a proxy for the entire stock market, and many investors and fund managers use the DJIA as a benchmark: if their investment strategies deliver better annual returns than the DJIA, this is seen as proof of their money-making abilities. A sure-fire strategy not just to predict but to beat the DJIA would therefore capture the attention of investors everywhere – and this is what Preis *et al.* claimed to be able to do. Using data for the period from January 2004 to February 2011, they found that certain keywords were predictive of DJIA performance, and they argued that investors could have profited handsomely by making buy or sell decisions based on changes in the search volume index scores of these keywords.

Their strategy (which I will call the "GT strategy") involved a simple moving average trading rule based on 98 keywords, which they used throughout the sample period. For each keyword each week, the researchers would buy into the DJIA if the keyword had trended downwards in the previous week, and they would sell the DJIA if the keyword had trended upwards. The following week, they would close off their trading positions by reversing the initial trade. This process was repeated for all weeks in the sample period, and at the end of the period the researchers calculated the return they made for each keyword and compared this to an alternative strategy, known as "buy and hold" (BH). The BH strategy simply bought into the DJIA index at the start of the sample period, and held this position until the end of the sample period.

Debt was the winning keyword in the Preis *et al.* study, producing a cumulative return of 326% over the sample period, leaving the BH strategy (with a 16% cumulative return) in the dust. That seems impressive – but the Preis *et al.* approach does present a couple of issues that can undermine the reliability of any big data analysis. First, there should be sound, principled arguments – ideally supported by evidence from prior studies – for the choice of predictive variables. Second, the predictive power of the chosen variables should be tested on data "held out" from the overall sample. Failure to meet these two conditions raises the risk of being fooled by fluke correlations that hold only for the sample under test and never again.

In the Preis *et al.* study, researchers did select keywords that were relevant to the concept of stock markets and finance. However, the decision of which keywords to use was not made during the sample period, but afterwards, so the selection may be tainted with hindsight bias. For example, the sample period covered by the Preis *et al.* study spans the global financial crisis of 2008–2009, so *debt* now seems like a natural choice of keyword, though that may not have been so obvious at the time. In addition, the list of keywords was fixed throughout the sample period, which seems like an unrealistic strategy for investors to adopt: those looking to predict stock market performance are perhaps more likely to choose keywords dynamically, adjusting their choice with changing market conditions. And, of course, the predictive power of the Preis *et al.* keywords was only tested on historical data, so it is not possible to say whether a trading strategy that *might* have worked in the past will *definitely* work in the present.

Another try

I addressed the above issues in a recent *Journal of Index Investing* paper.[6] I started with the same list of keywords as Preis *et al.*, but rather than applying the GT strategy to a fixed list for the entire period (2004–2011), I adopted a "rolling GT strategy", in which I ran a horse race among the 98 keywords

> The mouth-watering performance of my investment strategy is most likely spurious, given the lack of plausible economic stories for most of my best keywords

each year to determine the best keyword to use for predicting the DJIA the following year. For example, in year 1 of the sample period, I used the original GT strategy to determine the best keyword based on cumulative returns for that year. I then used the best keyword for year 1 as the basis of my investment strategy in year 2. Meanwhile, over the course of year 2, the horse race was rerun for the full list of 98 keywords, and again

the GT strategy was used to select the best keyword for year 2, which would then be the focus of my investment strategy in year 3, and so on. The performance of my rolling GT strategy would then be judged on the cumulative returns generated in years $X + 1$ by the best keywords from years X.

I do not claim that this method is optimal, but it did allow keywords to be chosen in an intuitive and dynamic way, much as they might in a real-world trading context, where past performance is used to inform future strategies. It also helped to avoid any look-ahead bias, which can be a problem when trading strategies are tested on historical data, as results may be skewed by information that would not have been known or available during the sample period.

My results showed firstly that *debt* was never the best keyword in any year. The best keywords for the years in chronological order were *society*, *cancer*, *home*, *holiday*, *inflation* and *color*. While one can argue that a high search volume for *debt* captures investors' nervousness and a low search volume for *debt* captures investors' optimism about the market's short-term trend, none of the above keywords (with the exception of *inflation*) have any economic content.

Rawpixel.com/bigstock.com

In terms of performance, my rolling GT strategy produced a cumulative return of 94%, compared to 13.5% for the BH strategy. This is a decent showing, but my 7-to-1 performance ratio is much lower than the 20-to-1 performance ratio reported by Preis *et al*. In any case, the mouth-watering performance of my strategy is most likely spurious, given the lack of plausible economic stories for most of my best keywords.

Spurious results will not survive out-of-sample tests. So, a natural extension is to repeat the previous experiment using a more recent period. In my journal paper, I chose the period from March 2011 to May 2016. For this experiment, I used a smaller set of 50 keywords, again drawn from the full Preis *et al.* list. As we have now seen, the full list included some words – such as *color*, *culture*, and *holiday* – which have no real economic relevance, so to avoid spurious results, all 50 keywords in this new list were directly related to the economy or financial markets. They included terms like *economy*, *recession*, *boom*, *Federal Reserve*, *interest rates*, *DJIA* (the index itself) and *S&P 500 index*. *Debt* was retained as a keyword in this list.

As before, I implemented the rolling GT strategy and compared its performance with the BH strategy. I will highlight two results of this out-of-sample test. First, once again, *debt* was never the best keyword. Second, unlike the 2004–2011 test, the rolling GT strategy performed dismally in this later period. A dollar invested in March 2011 was worth only $0.78 by May 2016, compared to $1.45 for the BH strategy. The lack of robustness of the rolling GT strategy should temper unwarranted optimism of Google Trends, specifically, as being a game changer for stock market prediction.

But hope springs eternal in the prediction business, given our hard-wired need to seek patterns. My study is not intended to dismiss the potential of big data in improving stock market prediction, but a reminder that attempts to monetise big data this way should be based on predictors that are backed by sound economic theory and out-of-sample forecasting reliability. Predictions that do not pass these tests are doomed to be misguided. ■

References

1. Ross, S. J. (2005) *Neoclassical Finance*. Princeton, NJ: Princeton University Press.

2. Goyal, A. and Welch, I. (2008) A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, **21**, 1455–1508.

3. Ang, A. and Bekaert, G. (2007) Stock return predictability: Is it there? *Review of Financial Studies*, **20**, 652–707.

4. Campbell, S. D. and Diebold, F. X. (2009) Stock returns and expected business conditions: Half a century of direct evidence. *Journal of Business and Economic Statistics*, **27**, 266–278.

5. Preis, T., Moat, H. S. and Stanley, H. E. (2013) Quantifying trading behavior in financial markets using *Google Trends*. *Scientific Reports*, **3**, art. no. 1684.

6. Fong, W. M. (2017) Big data, small pickings: Predicting the stock market using Google Trends. *Journal of Index Investing*, **7**, 75–82.