

Can Twitter sentiment predict stock market behaviour?



There are some positive indications, but this relationship is neither linear nor simple, writes [Francesco Corea](#).

The stock markets are an incredibly competitive environment and nowadays hedge funds and investment managers are rushing to find innovative ways to produce profits for their investors and their institutions. In addition to therefore understanding the impact of macroeconomic events or regulatory constraints on specific investments, they are looking at new datasets as main sources of competitive advantage.

Someone is using geo satellite images to gather data about the [parking lots at Wal-Mart stores](#) and predict their quarterly revenues; someone else is tracking down the number of drug tests sold to forecast the employee growth rate in specific companies and sectors; others are using [app check-in](#) to predict a loss in sales figures. All these examples are interesting but still small in scale, and not whatsoever comparable to the impact of social media in the financial services industry.

Many academics over the last decade have used similar datasets to try to anticipate trends in financial markets (using [stock message boards](#), [financial news](#), [blogs](#), [web search queries](#), etc.) but only more recent studies have exploited social media ([specifically microblogs](#)) for the same purpose. In particular, Twitter represents a gigantic source of information that, with the appropriate tools (namely, natural language processing techniques) can be used to infer the general sentiment around specific topics of interest.

This is indeed the main idea behind my work: what if having people complaining on Twitter about their malfunctioning iPhone will in fact result in Apple experience quarterly losses? Of course, the relationship is neither linear nor simple. It actually depends on the total number of users expressing a certain sentiment as well as on the intensity of their sentiments, and even in perfect circumstances, the correlation between general sentiment and stock oscillations needs to be proved.

I have therefore investigated this type of relationship for three major technology stocks (namely, Apple, Facebook, and Google). Even though my prior assumption was that this correlation might be actually stronger for smaller stocks, which are more volatile and easily influenced by the popular opinion, I needed a critical mass of tweets to create robust analysis on this phenomenon and this is why I preferred to analyse larger technology stocks.

I also decided to apply other two filters, which are the language of those tweets (I only selected English tweets) and I constrained the tweets that I wanted to analyse to those that showed a certain degree of financial literacy from the user (I selected only the ones that included the stock ticker in the 140 characters). This last step was essential to filter out irrelevant texts (“I am eating an apple and I do not like it”). In this way, I gathered over 160,000 tweets for a period of two months and assigned them a score swinging from -20 (extremely negative) to +20 (extremely positive).



In order to test the correlation between stocks and general sentiment, I then built a series of different indicators, some of them purely sentiment-driven, others with a stronger focus on the volume rather than the actual sentiment of the message, and others mixed, analyzing all of them at different frequencies. I also included the *Klout score* in the analysis, which is a value that indicates the degree of social influence of a certain individual in the social media world (it varies between 1 and 100 and to a higher value corresponds a higher influence power).

Even if the rationale of the work was finding patterns and common structures in popular sentiment that might impact the stock markets, I was not expecting all the indicators to be relevant so I used techniques that helped me automatically select the relevant features for this analysis.

The results were indeed quite unexpected: there was not a common structure across the different stocks that could have been identified. In other words, different stocks reacted to different indicators, probably because of some stock idiosyncrasies not previously considered.

Regardless of the specific indicator, though, both the sentiment of the tweets as well as their volumes were proven to have some predictive power for the financial markets when considered on a minute basis. It appeared also clear to me that the time span analysed was too short and that a longer time series is needed for this type of study to be robustly tested.

It also made me think that normal tweeting activity and trading days might be sentiment-resistant and that it could be more interesting to study this correlation when extraordinary corporate events happen (for example, an acquisition, an IPO, the launch of a new product, the company distributing dividends, the CEO been fired, etc.)

This study concludes highlighting the importance of social media data in trading strategies but, as often happens in academia, it eventually raises more questions than it provides answers. I think there is an incredible value in using alternative data in complex systems such as the financial markets, so I only hope this type of work will encourage others to experiment and test innovative ideas far from the traditional research path.

- This blog post is based on the author's paper [Can Twitter Proxy the Investors' Sentiment? The Case for the Technology Sector](#), Big Data Research, June 2016 and first appeared at the [LSE Business Review](#).
- Featured image credit: [Twitter](#) (cropped), by [@G_Obieta](#) for [Rosaura Ochoa](#), under a [CC-BY-2.0](#) licence

[Please read our comments policy before commenting.](#)

Note: This article gives the views of the author, and not the position of USAPP– American Politics and Policy, nor of the London School of Economics.

Shortened URL for this post: <http://bit.ly/2yjjvKH>

About the author



Francesco Corea

Francesco Corea is a complexity scientist, AI entrepreneur and tech investor, and he runs the blog [Cyber Tales](#). Francesco is a strong supporter of an interdisciplinary research approach, and he wants to foster the interaction of different sciences in order to bring to light hidden connections. He is a former Anthemis Fellow, IPAM Fellow, and he got his PhD from LUISS University. His topics of interests are big data and AI, and he focuses on fintech, medtech, and energy verticals.