OPEN

Quantifying Trading Behavior in Financial Markets Using *Google Trends*

Tobias Preis[1]*, Helen Susannah Moat[2,3]* & H. Eugene Stanley[2]*

[1]Warwick Business School, University of Warwick, Scarman Road, Coventry, CV4 7AL, UK, [2]Department of Physics, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215, USA, [3]Department of Civil, Environmental and Geomatic Engineering, UCL, Gower Street, London, WC1E 6BT, UK.

Crises in financial markets affect humans worldwide. Detailed market data on trading decisions reflect some of the complex human behavior that has led to these crises. We suggest that massive new data sources resulting from human interaction with the Internet may offer a new perspective on the behavior of market participants in periods of large market movements. By analyzing changes in *Google* query volumes for search terms related to finance, we find patterns that may be interpreted as "early warning signs" of stock market moves. Our results illustrate the potential that combining extensive behavioral data sets offers for a better understanding of collective human behavior.

The increasing volumes of 'big data' reflecting various aspects of our everyday activities represent a vital new opportunity for scientists to address fundamental questions about the complex world we inhabit[1–7]. Financial markets are a prime target for such quantitative investigations[8,9]. Movements in the markets exert immense impacts on personal fortunes and geopolitical events, generating considerable scientific attention to this subject[10–19]. For example, a range of recent studies have focused on modeling financial markets[20–25] and on performing network analyses[26–29].

At their core, financial trading data sets reflect the myriad of decisions taken by market participants. According to Herbert Simon, actors begin their decision making processes by attempting to gather information[30]. In today's world, information gathering often consists of searching online sources. Recently, the search engine *Google* has begun to provide access to aggregated information on the volume of queries for different search terms and how these volumes change over time, via the publicly available service *Google Trends*. In the present study, we investigate the intriguing possibility of analyzing search query data from *Google Trends* to provide new insights into the information gathering process that precedes the trading decisions recorded in the stock market data.

A recent investigation has shown that the number of clicks on search results stemming from a given country correlates with the amount of investment in that country[31]. Further studies exploiting the temporal dimension of *Google Trends* data have demonstrated that changes in query volumes for selected search terms mirror changes in current numbers of influenza cases[32] and current volumes of stock market transactions[33]. This demonstration of a link between stock market transaction volume and search volume has also been replicated using *Yahoo!* data[34]. Choi and Varian[35] have shown that data from *Google Trends* can be linked to current values of various economic indicators, including automobile sales, unemployment claims, travel destination planning and consumer confidence. A very recent study has shown that Internet users from countries with a higher per capita GDP are more likely to search for information about years in the future than years in the past[36].

Here, we suggest that within the time period we investigate, *Google Trends* data did not only reflect the current state of the stock markets[33] but may have also been able to anticipate certain future trends. Our findings are consistent with the intriguing proposal that notable drops in the financial market are preceded by periods of investor concern. In such periods, investors may search for more information about the market, before eventually deciding to buy or sell. Our results suggest that, following this logic, during the period 2004 to 2011 *Google Trends* search query volumes for certain terms could have been used in the construction of profitable trading strategies.

Results

We analyze the performance of a set of 98 search terms. We included terms related to the concept of stock markets, with some terms suggested by the *Google Sets* service, a tool which identifies semantically related keywords. The set of terms used was therefore not arbitrarily chosen, as we intentionally introduced some financial bias. We explain our strategy based on changes in search volume with reference to the term *debt*, a
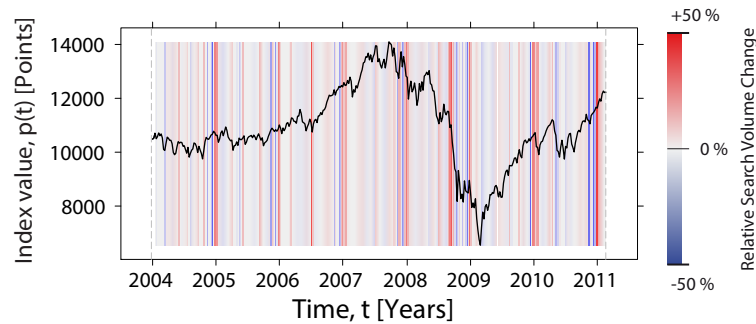
Figure 1 | Search volume data and stock market moves. Time series of closing prices $p(t)$ of the *Dow Jones Industrial Average* (DJIA) on the first day of trading in each week t covering the period from 5 January 2004 until 22 February 2011. The color code corresponds to the relative search volume changes for the search term *debt*, with $\Delta t = 3$ weeks. Search volume data are restricted to requests of users localized in the United States of America.

keyword with an obvious semantic connection to the most recent financial crisis, and overall the term which performed best in our analyses.

To uncover the relationship between the volume of search queries for a specific term and the overall direction of trader decisions, we analyze closing prices $p(t)$ of the *Dow Jones Industrial Average* (DJIA) on the first trading day of week t. We use *Google Trends* to determine how many searches $n(t-1)$ have been carried out for a specific search term such as *debt* in week $t-1$, where *Google* defines weeks as ending on a Sunday, relative to the total number of searches carried out on *Google* during that time. We find that search volume data change slightly over time due to *Google*'s extraction procedure. For each search term, we therefore average over three realizations of its search volume time series, based on three independent data requests in consecutive weeks. The variability of *Google Trends* data across different dates of access is irrelevant for our results, and it can be shown that the data are consistent with reported real world events (see Fig. S1 in the *Supplementary Information*).

To quantify changes in information gathering behavior, we use the relative change in search volume: $\Delta n(t, \Delta t) = n(t) - N(t-1, \Delta t)$ with $N(t-1, \Delta t) = (n(t-1) + n(t-2) + \ldots + n(t-\Delta t))/\Delta t$, where t is measured in units of weeks. In Fig. 1, we depict relative search volume changes for the term *debt*, and their relationship to DJIA closing prices.

To investigate whether changes in information gathering behavior as captured by *Google Trends* data were related to later changes in stock price in the period between 2004–2011, we implement a hypothetical investment strategy for a portfolio using search volume data, called '*Google Trends* strategy' in the following. Profit can only be made in a trading strategy if at least some future changes in the stock price are correctly anticipated, in particular around large market

movements. We implement this strategy by selling the DJIA at the closing price $p(t)$ on the first trading day of week t, if $\Delta n(t-1, \Delta t) > 0$, and buying the DJIA at price $p(t+1)$ at the end of the first trading day of the following week. Note that mechanisms exist which make it possible to sell assets in financial markets without first owning them. If instead $\Delta n(t-1, \Delta t) < 0$, then we buy the DJIA at the closing price $p(t)$ on the first trading day of week t and sell the DJIA at price $p(t+1)$ at the end of the first trading day of the coming week. At the beginning of trading, we set the value of all portfolios to an arbitrary value of 1. If we take a 'short position'—selling at the closing price $p(t)$ and buying back at price $p(t+1)$—then the cumulative return R changes by $\log(p(t)) - \log(p(t+1))$. If we take a 'long position'—buying at the closing price $p(t)$ and selling at price $p(t+1)$—then the cumulative return R changes by $\log(p(t+1)) - \log(p(t))$. In this way, buy and sell actions have symmetric impacts on the cumulative return R of a strategy's portfolio. In using this approach to analyze the relationship between *Google* search volume and stock market movements, we neglect transaction fees, since the maximum number of transactions per year when using our strategy is only 104, allowing a closing and an opening transaction per week. We of course do not dispute that such transaction fees would impact profit in a real world implementation.

In Fig. 2, the performance of the *Google Trends* strategy based on the search term *debt* is depicted by a blue line, whereas dashed lines indicate the standard deviation of the cumulative return from a strategy in which we buy and sell the market index in an uncorrelated, random manner ('random investment strategy'). The standard deviation is derived from simulations of 10,000 independent realizations of the random investment strategy. Fig. 2 shows that the use of the *Google Trends* strategy, based on the search term *debt* and $\Delta t = 3$ weeks, would have increased the value of a portfolio by 326%. The
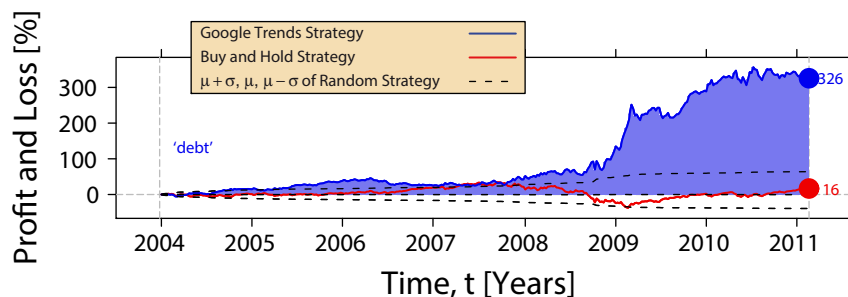

Figure 2 | Cumulative performance of an investment strategy based on *Google Trends* data. Profit and loss for an investment strategy based on the volume of the search term *debt*, the best performing keyword in our analysis, with $\Delta t = 3$ weeks, plotted as a function of time (*blue line*). This is compared to the "buy and hold" strategy (*red line*) and the standard deviation of 10,000 simulations using a purely random investment strategy (*dashed lines*). The *Google Trends* strategy using the search volume of the term *debt* would have yielded a profit of 326%.
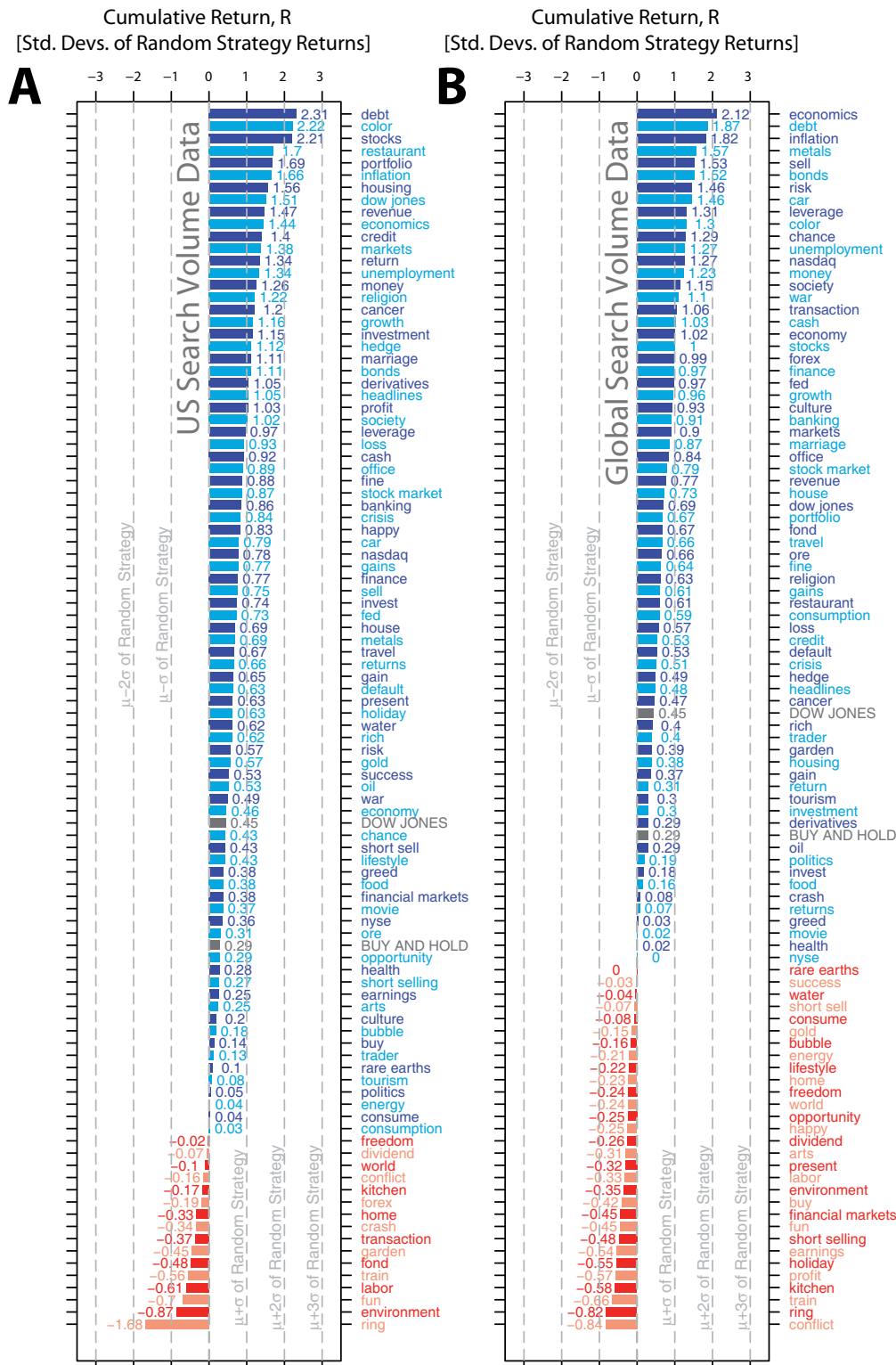
Figure 3 | Performances of investment strategies based on search volume data. (**A**) Cumulative returns of 98 investment strategies based on search volumes restricted to search requests of users located in the United States for different search terms, displayed for the entire time period of our study from 5 January 2004 until 22 February 2011—the time period for which *Google Trends* provides data. We use two shades of blue for positive returns and two shades of red for negative returns to improve the readability of the search terms. The cumulative performance for the "buy and hold strategy" is also shown, as is a "Dow Jones strategy", which uses weekly closing prices of the *Dow Jones Industrial Average* (DJIA) rather than *Google Trends* data (see gray bars). Figures provided next to the bars indicate the returns of a strategy, R, in standard deviations from the mean return of uncorrelated random investment strategies, $<R>_{RandomStrategy} = 0$. Dashed lines correspond to -3, -2, -1, 0, $+1$, $+2$, and $+3$ standard deviations of random strategies. We find that returns from the *Google Trends* strategies tested are significantly higher overall than returns from the random strategies ($<R>_{US} = 0.60$; $t = 8.65$, $df = 97$, $p < 0.001$, one sample t-test). (**B**) A parallel analysis shows that extending the range of the search volume analysis to global users reduces the overall return achieved by *Google Trends* trading strategies on the U.S. market ($<R>_{US} = 0.60$, $<R>_{Global} = 0.43$; $t = 2.69$, $df = 97$, $p < 0.01$, two-sided paired t-test). However, returns are still significantly higher than the mean return of random investment strategies ($<R>_{Global} = 0.43$; $t = 6.40$, $df = 97$, $p < 0.001$, one sample t-test).
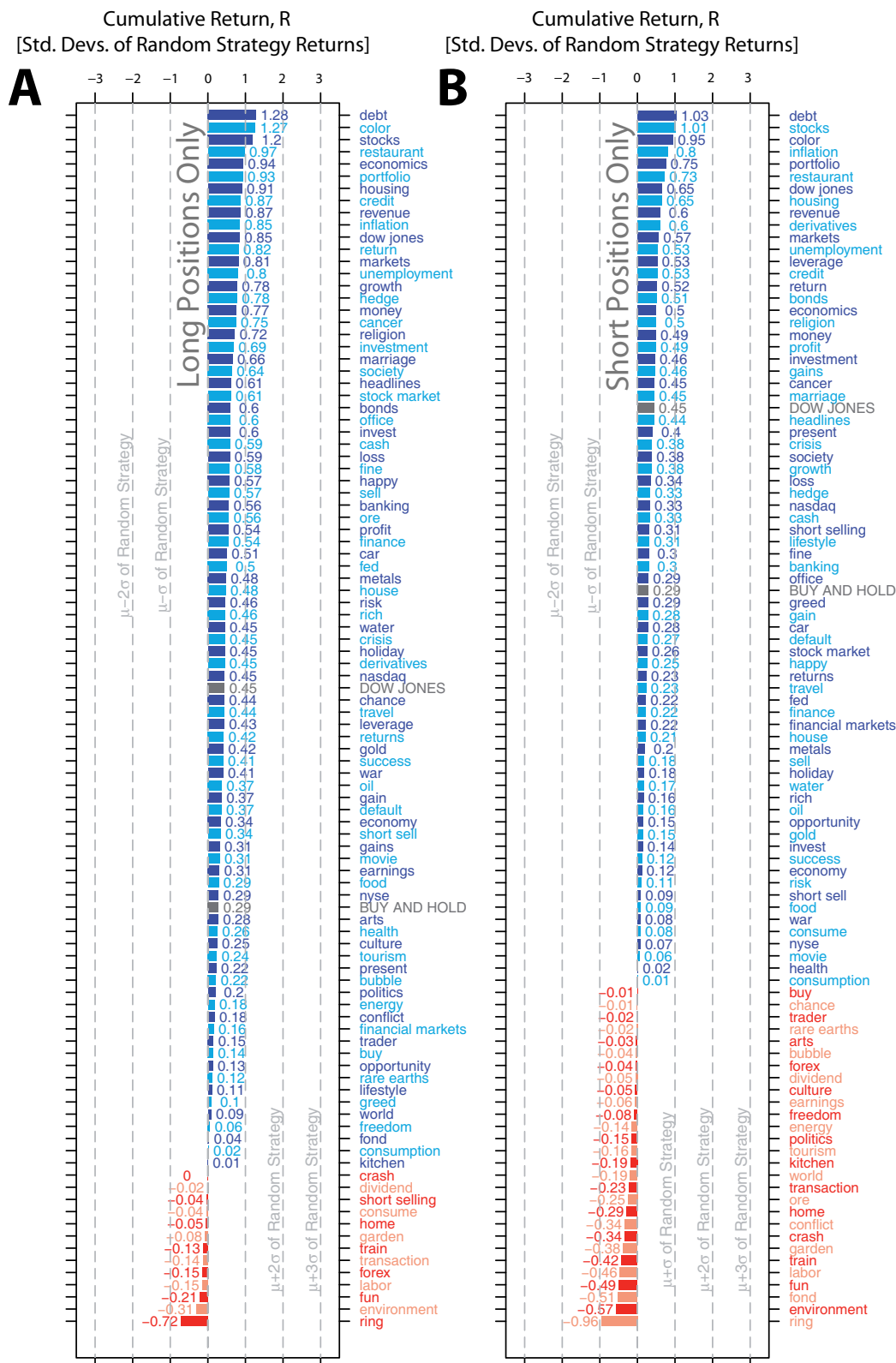
Figure 4 | Analysis using strategies in which we take long or short positions only, using U.S. search volume data. (**A**) We implement *Google Trends* strategies in which we take long positions following a decrease in search volume, and never take short positions. We find that returns from these long position *Google Trends* strategies are significantly higher overall than returns from the random investment strategies ($<R>_{USLong} = 0.41$; $t = 11.42$, $df = 97$, $p < 0.001$, one sample t-test). Again, we find a positive correlation between our indicator of financial relevance and returns from these strategies (Kendall's tau = 0.242, $z = 3.53$, $N = 98$, $p < 0.001$). (**B**) We also implement *Google Trends* strategies in which we take short positions following an increase in search volume, and never take long positions. In line with our results from the long position *Google Trends* strategies, we find that returns from the short position *Google Trends* strategies are significantly higher overall than returns from the random investment strategies ($<R>_{USShort} = 0.19$; $t = 5.28$, $df = 97$, $p < 0.001$, one sample t-test), and that there is a positive correlation between our indicator of financial relevance and short position *Google Trends* returns (Kendall's tau = 0.275, $z = 4.01$, $N = 98$, $p < 0.001$).

performance of *Google Trends* strategies based on all other search terms that we analyze is depicted in Figures S3-S100 in the *Supplementary Information*.

We rank the full list of the 98 investigated search terms by their trading performance when using search data for U.S. users only (Fig. 3A) and when using globally generated search volume (Fig. 3B). In order to ensure the robustness of our results, the overall performance of a strategy based on a given search term is determined as the mean value over the six returns obtained for $\Delta t = 1 \ldots 6$ weeks. Returns of the strategies are calculated as the logarithm of relative portfolio changes, following the usual definition of returns. The distribution of final portfolio values resulting from the random investment strategies is close to log-normal. Cumulative returns from the random investment strategy, derived from the logarithm of these portfolio values, therefore follow a normal distribution, with a mean value of $<R>_{RandomStrategy} = 0$. Here we report R, the cumulative returns of a strategy, in standard deviations of the cumulative returns of these uncorrelated random investment strategies.

We find that returns from the *Google Trends* strategies we tested are significantly higher overall than returns from the random strategies ($<R>_{US} = 0.60$; $t = 8.65$, $df = 97$, $p < 0.001$, one sample t-test).

We compare the performance of these search terms with two benchmark strategies. The 'buy and hold' strategy is implemented by buying the index in the beginning and selling it at the end of the hold period. This strategy yields 16% profit, equal to the overall increase in value of the DJIA in the time period from January 2004 until February 2011. We further implement a '*Dow Jones* strategy' by using changes in $p(t)$ in place of changes in search volume data as the basis of buy and sell decisions. We find that this strategy also yields only 33% profit with $\Delta t = 3$ weeks, or when determined as the mean value over the six returns obtained for $\Delta t = 1 \ldots 6$ weeks, 0.45 standard deviations of cumulative returns of uncorrelated random investment strategies (Figs. 3A and 3B; see also Fig. S101 in the *Supplementary Information*).

Our results show that performance of the *Google Trends* strategy differs with the search term chosen. We investigate whether these differences in performance can be partially explained using an indicator of the extent to which different terms are of financial relevance—a concept we quantify by calculating the frequency of each search term in the online edition of the *Financial Times* from August 2004 to June 2011, normalized by the number of *Google* hits for each search term (see Fig. S2 in the Supplementary Information). We find that the return associated with a given search term is correlated with this indicator of financial relevance (Kendall's tau = 0.275, $z = 4.01$, $N = 98$, $p < 0.001$) using Kendall's tau rank correlation coefficient[37].

It is widely recognized that investors prefer to trade on their domestic market, suggesting that search data for U.S. users only, as used in analyses so far, should better capture the information gathering behavior of U.S. stock market participants than data for *Google* users worldwide. Indeed, we find that strategies based on global search volume data are less successful than strategies based on U.S. search volume data in anticipating movements of the U.S. market ($<R>_{US} = 0.60$, $<R>_{Global} = 0.43$; $t = 2.69$, $df = 97$, $p < 0.01$, two-sided paired t-test).

Our empirical results so far are consistent with a two part hypothesis: namely that key increases in the price of the DJIA were preceded by a decrease in search volume for certain financially related terms, and conversely, that key decreases in the price of the DJIA were preceded by an increase in search volume for certain financially related terms. However, our trading strategy can be decomposed into two strategy components: one in which a decrease in search volume prompts us to buy (or take a long position) and one in which an increase in search volume prompts us to sell (or take a short position).

In order to verify that both strategy components play a significant role in our results, such that we have evidence for both parts of this hypothesis, we implement and test one strategy in which we take long positions following a decrease in search volume but never take short positions (Fig. 4A), and another strategy in which we take short positions following an increase in search volume but never take long positions (Fig. 4B). We find that returns from both *Google Trends* strategy components are significantly higher overall than returns from a random investment strategy (long position strategies: $<R>_{USLong} = 0.41$; $t = 11.42$, $df = 97$, $p < 0.001$, one sample t-test; short position strategies: $<R>_{USShort} = 0.19$; $t = 5.28$, $df = 97$, $p < 0.001$, one sample t-test).

Discussion

In summary, our results are consistent with the suggestion that during the period we investigate, *Google Trends* data did not only reflect aspects of the current state of the economy, but may have also provided some insight into future trends in the behavior of economic actors. Using historic data from the period between January 2004 and February 2011, we detect increases in *Google* search volumes for keywords relating to financial markets before stock market falls. Our results suggest that these warning signs in search volume data could have been exploited in the construction of profitable trading strategies.

We offer one possible interpretation of our results within the context of Herbert Simon's model of decision making[28]. We suggest that *Google Trends* data and stock market data may reflect two subsequent stages in the decision making process of investors. Trends to sell on the financial market at lower prices may be preceded by periods of concern. During such periods of concern, people may tend to gather more information about the state of the market. It is conceivable that such behavior may have historically been reflected by increased *Google Trends* search volumes for terms of higher financial relevance.

We find that strategies based on search volume data for U.S. users are more successful for the U.S. market than strategies using global search volume data. Given the assumption that the population of U.S. Internet users contains a higher proportion of traders on the U.S. markets than the worldwide population of Internet users contains, this finding is in line with the intriguing suggestion that these datasets may provide insights into different stages of decision making within the same population.

In this work, we provide a quantification of the relationship between changes in search volume and changes in stock market prices. Future work will be needed to provide a thorough explanation of the underlying psychological mechanisms which lead people to search for terms like *debt* before selling stocks at a lower price. It is clear that many opportunities also remain to extend our analyses to further financial data sets.

The results of our investigation suggest that combining large behavioral data sets such as financial trading data with data on search query volumes may open up new insights into different stages of large-scale collective decision making. We conclude that these results further illustrate the exciting possibilities offered by new big data sets to advance our understanding of complex collective behavior in our society.

Methods

How related are search terms to the topic of finance? We quantify financial relevance by calculating the frequency of each search term in the online edition of the *Financial Times* (http://www.ft.com) from August 2004 to June 2011, normalized by the number of *Google* hits (http://www.google.com) for each search term. Details are given in the *Supplementary Information*.

Data retrieval. We retrieved search volume data by accessing the *Google Trends* website (http://www.google.com/trends) on 10 April 2011, 17 April 2011, and 24 April 2011. The data on the number of hits for search terms in the online edition of the *Financial Times* was retrieved on 7 June 2011. The numbers of *Google* hits for these terms were obtained on 8 June 2011.

1. Axtell, R. L. Zipf distribution of US firm sizes. *Science* **293**, 1818–1820 (2001).
2. King, G. Ensuring the Data-Rich Future of the Social Sciences. *Science* **331**, 719–721 (2011).

3. Vespignani, A. Predicting the Behavior of Techno-Social Systems. *Science* **325**, 425–428 (2009).
4. Lazer, D. *et al.* Computational Social Science. *Science* **323**, 721–723. (2009).
5. Perc, M. Evolution of the most common English words and phrases over the centuries. *J. R. Soc. Interface* **9**, 3323–3328 (2012).
6. Petersen, A. M., Tenenbaum, J. N., Havlin, S., Stanley, H. E. & Perc, M. Languages cool as they expand: Allometric scaling and the decreasing need for new words. *Scientific Reports* **2**, 943 (2012).
7. Christakis, N. A. & Fowler, J. H. *Connected: The surprising power of our social networks and how they shape our lives* (Little, Brown and Company, 2009).
8. Fehr, E. Behavioural science - The economics of impatience. *Nature* **415**, 269–272 (2002).
9. Shleifer, A. *Inefficient Markets: An Introduction to Behavioral Finance* (Oxford University Press, Oxford, 2000).
10. Lillo, F., Farmer, J. D. & Mantegna, R. N. Econophysics - Master curve for price-impact function. *Nature* **421**, 129–130 (2003).
11. Gabaix, X., Gopikrishnan, P., Plerou, V. & Stanley, H. E. A theory of power-law distributions in financial market fluctuations. *Nature* **423**, 267–270 (2003).
12. Preis, T., Kenett, D. Y., Stanley, H. E., Helbing, D. & Ben-Jacob, E. Quantifying the Behavior of Stock Correlations Under Market Stress. *Scientific Reports* **2**, 752 (2012).
13. Preis, T., Schneider, J. J. & Stanley, H. E. Switching processes in financial markets. *PNAS* **108**, 7674–7678 (2011).
14. Preis, T. Econophysics - complex correlations and trend switchings in financial time series. *European Physical Journal Special Topics* **194**, 5–86 (2011).
15. Bunde, A., Schellnhuber, H. J. & Kropp, J., eds. *The Science of Disasters: Climate Disruptions; Heart Attacks; and Market Crashes* (Springer, Berlin, 2002).
16. Vandewalle, N. & Ausloos, M. Coherent and random sequences in financial fluctuations. *Physica A* **246**, 454–459 (1997).
17. Podobnik, B., Horvatic, D., Petersen, A. M. & Stanley, H. E. Cross-correlations between volume change and price change. *PNAS* **106**, 22079–22084 (2009).
18. Sornette, D., Woodard, R. & Zhou, W. X. The 2006-2008 oil bubble: Evidence of speculation, and prediction. *Physica A* **388**, 1571–1576. (2009).
19. Watanabe, K., Takayasu, H. & Takayasu, M. A mathematical definition of the financial bubbles and crashes. *Physica A* **383**, 120–124 (2007).
20. Bouchaud, J. P., Matacz, A. & Potters, M. Leverage effect in financial markets: the retarded volatility model. *Physical Review Letters* **87**, 228701 (2001).
21. Hommes, C. H. Modeling the stylized facts in finance through simple nonlinear adaptive systems. *PNAS* **99**, 7221–7228 (2002).
22. Haldane, A. G. & May, R. M. Systemic risk in banking ecosystems. *Nature* **469**, 351–355 (2011).
23. Lux, T. & Marchesi, M. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature* **397**, 498–500 (1999).
24. Krugman, P. *The Self-Organizing Economy* (Blackwell, Cambridge, Massachusetts, 1996).
25. Sornette, D. & von der Becke, S. Complexity clouds finance-risk models. *Nature* **471**, 166 (2011).
26. Schweitzer, F. *et al.* Economic Networks: The New Challenges. *Science* **325**, 422–425 (2009).
27. Garlaschelli, D., Caldarelli, G. & Pietronero, L. Universal scaling relations in food webs. *Nature* **423**, 165–168 (2003).
28. Onnela, J. P., Arbesman, S., Gonzalez, M. C., Barabasi, A. L. & Christakis, N. A. Geographic Constraints on Social Network Groups. *PLoS One* **6**, e16939 (2011).
29. Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E. & Havlin, S. Catastrophic cascade of failures in interdependent networks. *Nature* **464**, 1025–1028 (2010).
30. Simon, H. A. A behavioral model of rational choice. *Quarterly Journal of Economics* **69**, 99–118 (1955).
31. Mondria, J., Wu, T. & Zhang, Y. The determinants of international investment and attention allocation: Using internet search query data. *Journal of International Economics* **82**, 85–95 (2010).
32. Ginsberg, J. *et al.* Detecting influenza epidemics using search engine query data. *Nature* **457**, 1012–1014 (2009).
33. Preis, T., Reith, D. & Stanley, H. E. Complex dynamics of our economic life on different scales: insights from search engine query data. *Phil. Trans. R. Soc. A* **368**, 5707–5719 (2010).
34. Bordino, I. *et al.* Web Search Queries Can Predict Stock Market Volumes. *PLoS One* **7**, e40014 (2012).
35. Choi, H. & Varian, H. Predicting the Present with Google Trends. *The Economic Record* **88**, 2–9 (2012).
36. Preis, T., Moat, H. S., Stanley, H. E. & Bishop, S. R. Quantifying the Advantage of Looking Forward. *Scientific Reports* **2**, 350 (2012).
37. Kendall, M. A New Measure of Rank Correlation. *Biometrika* **30**, 81–89 (1938).

Acknowledgements

Author contributions

T.P., H.S.M. and H.E.S. developed the design of the study, performed analyses, discussed the results, and contributed to the text of the manuscript.

Additional information

How to cite this article: Preis, T., Moat, H.S. & Stanley, H.E. Quantifying Trading Behavior in Financial Markets Using *Google Trends*. *Sci. Rep.* **3**, 1684; DOI:10.1038/srep01684 (2013).