Downloaded from http://rsos.royalsocietypublishing.org/ on September 15, 2016

ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Research

Cite this article: Garcia D, Schweitzer F. 2015 Social signals and algorithmic trading of Bitcoin. *R. Soc. open sci.* **2**: 150288.
http://dx.doi.org/10.1098/rsos.150288

Received: 23 June 2015
Accepted: 28 August 2015

Subject Category:
Computer science

Subject Areas:
e-science

Keywords:
Bitcoin, computational social science, algorithmic trading, polarization, sentiment, prediction

Author for correspondence:
David Garcia
e-mail: dgarcia@ethz.ch

Electronic supplementary material is available at http://dx.doi.org/10.1098/rsos.150288 or via http://rsos.royalsocietypublishing.org.

THE ROYAL SOCIETY PUBLISHING

Social signals and algorithmic trading of Bitcoin

David Garcia and Frank Schweitzer

Chair of Systems Design, ETH Zurich, Weinbergstrasse 56/58, Zurich 8092, Switzerland

The availability of data on digital traces is growing to unprecedented sizes, but inferring actionable knowledge from large-scale data is far from being trivial. This is especially important for computational finance, where digital traces of human behaviour offer a great potential to drive trading strategies. We contribute to this by providing a consistent approach that integrates various datasources in the design of algorithmic traders. This allows us to derive insights into the principles behind the profitability of our trading strategies. We illustrate our approach through the analysis of Bitcoin, a cryptocurrency known for its large price fluctuations. In our analysis, we include economic signals of volume and price of exchange for USD, adoption of the Bitcoin technology and transaction volume of Bitcoin. We add social signals related to information search, word of mouth volume, emotional valence and opinion polarization as expressed in tweets related to Bitcoin for more than 3 years. Our analysis reveals that increases in opinion polarization and exchange volume precede rising Bitcoin prices, and that emotional valence precedes opinion polarization and rising exchange volumes. We apply these insights to design algorithmic trading strategies for Bitcoin, reaching very high profits in less than a year. We verify this high profitability with robust statistical methods that take into account risk and trading costs, confirming the long-standing hypothesis that trading-based social media sentiment has the potential to yield positive returns on investment.

1. Introduction

Our online society generates data on the digital traces of human behaviour at unprecedented scales and resolutions. This produces a *data deluge*, in which researchers are confronted with a vast amount of observational data that is not the product of carefully designed experiments [1]. One of the main challenges of the scientific community is to develop methods to extract meaningful knowledge from that data beyond mere descriptive analyses [2]. This is particularly important in financial trading: data can

© 2015 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.

Downloaded from http://rsos.royalsocietypublishing.org/ on September 15, 2016

be available to all financial agents, but it is the analysis and its applications which makes a difference. Within computational finance, the field of algorithmic trading [3] deals with the implementation and evaluation of automatic trading strategies, which are often kept in private companies and away from publicly accessible research. The most common kind of algorithmic trading is based on the principles of *technical analysis* [4], using the time series of prices to formulate predictions about returns. Technical analysis is often insufficient to derive satisfactory returns [5], motivating the inclusion of large-scale social signals and the evaluation through data-driven simulations on historical data, called *backtesting* [6,7]. In this article, we present a set of methods to derive stylized facts from the analysis of multidimensional economic and social signals, and to apply that knowledge in the design and evaluation of algorithmic trading strategies. We illustrate an application of our approach to algorithmic trading of the Bitcoin cryptocurrency, using a wide variety of digital traces about economic and social aspects of the Bitcoin ecosystem.

Bitcoin (BTC) is a digital currency designed to operate in a distributed system without any central authority, based on a cryptographic protocol that does not require a trusted third party [8]. Introduced in a 2008 paper written under the pseudonym of Satoshi Nakamoto [9], Bitcoin serves as a technology to transfer money quickly for negligible fees [10]. One of the first markets to adopt Bitcoin was the *Silk Road*, a website where illegal commerce became possible thanks to the relative anonymity of Bitcoin [11], in line with the evidence in search trends that relates Bitcoin usage to computer expertise and illegal activities [12]. Since then, the use of Bitcoin has widely expanded beyond criminal activities: at the time of writing, Bitcoin is accepted by many legal merchants and charities [13], including large businesses like Dell [14]. Bitcoin-accepting businesses, exchange markets and wallet services compose the *Bitcoin ecosystem* [8], where different kinds of agents interact, trade and communicate through digital channels. The increasing adoption of Bitcoin and its online nature allow us to simultaneously monitor its social and economic aspects. Every purchase of goods or services in Bitcoin leaves a trace in a public ledger called the *Block Chain*, creating a publicly accessible economic network [15]. Bitcoin's delocalized technology aligns with the online interaction of its users through social networks and forums, motivating its adoption by new users through word-of-mouth [16]. Previous research has shown how search trends and Wikipedia views are related to price changes [17] and to the speculative and monetary aspects of Bitcoin [18], leading to dynamics that combine search interest, user adoption, word-of-mouth and prices [16].

1.1. Contributions of this article

Based on established principles of time-series analysis and financial trading, we present a framework to derive general knowledge from multidimensional data on social and economic aspects of a market. We apply a general statistical model to detect temporal patterns in the co-movement of price and other signals. Those patterns are tested through a method robust to the empirical properties of the analysed data, formulating concise principles on which signals precede market movements. We combine those principles to produce tractable trading strategies, which we evaluate over a leave-out sample of the data, quantifying their profitability. Our approach, rather than focusing on improving a particular method, takes a multidisciplinary stance in which we combine principles from social psychology and economics with methods from information retrieval, time-series analysis and computational finance.

We apply our framework to the Bitcoin ecosystem, monitoring the digital traces of Bitcoin users with daily resolution. We combine *economic signals* related to market growth, trading volume, and use of Bitcoin as means of exchange, with *social signals* including search volumes, word-of-mouth levels, emotional valence and opinion polarization about Bitcoin. Our results reveal which signals precede changes of Bitcoin prices, a knowledge that we use to design algorithmic trading strategies. We evaluate the power of our strategies through backtesting data-driven simulations, comparing returns with technical analysis strategies. As a consequence, we test the hypothesis that social media sentiment predicts financial returns in the Bitcoin ecosystem.

1.2. Social signals in finance

Understanding the role of social signals in finance not only has the potential to generate significant profits, but also has scientific relevance as a research question [19]. Two different research approaches give insights to this question: one is the *statistical analysis* of social and financial signals in order to test the existence of temporal correlations that lead financial markets. The second one applies these signals in *prediction scenarios*, measuring their accuracy as a validation of the underlying behaviour of the system,

Downloaded from http://rsos.royalsocietypublishing.org/ on September 15, 2016

but not necessarily of their profitability. The statistical analysis of search engine data reveals that search trends can predict trading volumes of individual stocks [20]. In addition, stock prices in S&P 500 are correlated with tweet volumes [21], but the applicability of these patterns into trading strategies is yet to be evaluated.

Sentiment in social media is closely related to socio-economic phenomena, including public opinion [22]. This motivates the application of sentiment indicators in the statistical analysis of financial data. Early works on the sentiment in specialized forums gave negative results about their impact on returns [23]. Further research showed that emotions in private instant messaging between workers of a trading company precede stages of market volatility [24]. The expression of anxiety in publicly accessible data from general blogs precedes trading peaks and price drops in the S&P 500 [25], and sentiment in Twitter can be used to predict movements in large-scale stock indices [26]. It is important to note that, to date, there is no evidence that such sentiment-based predictions produce significant returns on investment [19].

1.3. Online polarization

While most previous works on sentiment in financial markets focus on dimensions of valence or mood, the collective phenomenon of polarization of opinions is often overlooked. The emergence of polarization in a society gives early warnings on political and economic phenomena: polarization in social networks of Swiss politicians precedes controversial elections [27], and polarization patterns in the Eurovision Song Contest appear before states of distrust in the European economy [28]. With respect to financial markets, speculation theories point to the role of diverse beliefs in financial transactions [29], leading to the hypothesis that polarization and disagreement influence trading volumes and prices [30]. In this line, the empirical analysis of polarization in stock message boards shows that states of disagreement lead to increased volatility [31].

1.4. The missing link

To date, there is a significant knowledge gap between the analysis and application of social signals to trading scenarios. Findings from statistical analyses alone are not guaranteed to lead to profitable strategies at all [25]. For example, movements of the Dow Jones Industrial Average (DJIA) can be predicted with mass media sentiment [32] and Twitter mood [26], but to date no research has shown that such prediction methods can be profitable in trading scenarios. Similarly, the analysis of discussion patterns in specialized blogs predict returns of some technology companies [33], but it is still open to evaluate the potential returns of such a predictor. The application of methods that process arbitrarily large datasets lead to results difficult to apply, for example the predicting power of search volumes of the query *'moon patrol'* [34] in backtesting over the DJIA [6]. Furthermore, analyses of Twitter discussions about companies can be applied in a portfolio strategy, yet its evaluation through backtesting leads to very moderate returns and their statistical significance is not assessed [35]. In addition, no previous research has proposed a prediction technique that derives significant returns on investment from online sentiment data [19]. Our research aims at closing the gap between these lines of research. To do so, we unify the statistical analysis and its application to design and evaluate trading strategies, based on tractable principles with potential impact in the finance community.

2. Trading strategy framework

To design and evaluate trading strategies, we present a framework that uses a set of economic and social signals related to the agents of the market under scrutiny. Among those signals, the only required one is an economic signal of prices of an asset, namely a stock, currency, or tradable index. To understand profitability, we convert the price time-series $P(t)$ into a return time series:

$$\mathrm{Ret}(t) = \frac{P(t) - P(t-1)}{P(t-1)}, \tag{2.1}$$

which quantifies proportional changes in the price at every time step. The data on these signals are divided in an analysis period and a leave-out period, as depicted in figure 1. The division in these periods needs to allocate enough data in the leave-out sample to provide the testing power to assess the statistical significance of strategy profits. For daily trading, a leave-out period of about 1 year is usually sufficient, but this ultimately depends on the expected profitability and variance of the trading strategies.

Downloaded from http://rsos.royalsocietypublishing.org/ on September 15, 2016

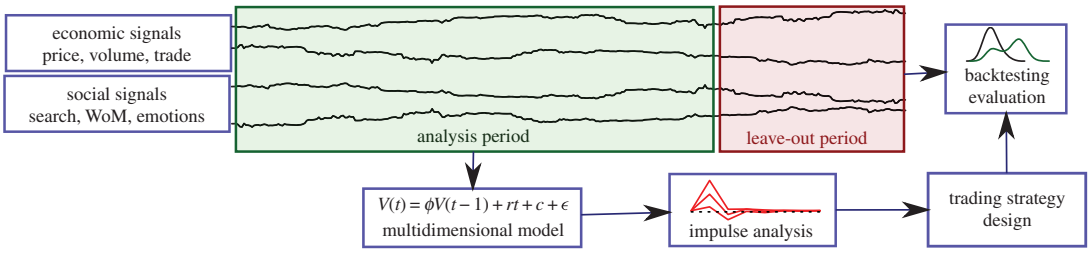

Figure 1. Framework for analysis of social and economic signals and trading strategy design and evaluation.

2.1. Multidimensional analysis

The first step in our framework focuses on the analysis period, applying a multidimensional model of vector auto-regression (VAR) [36], which is commonly used in the analysis of multidimensional time series in finance [16,23,37]. A VAR models multidimensional linear relations with given lags, which in our analysis we set to 1 day. Thus, given the vector of signals $V(t)$, we fit the equation

$$V(t) = \phi V(t-1) + r \times t + c + \epsilon, \tag{2.2}$$

where ϕ is a matrix of weights of the linear relations between variables, r is a deterministic trend vector, c is the vector of constant intercepts and ϵ is a vector of uncorrelated errors. While more advanced models can be considered, including longer lags and nonlinear terms, we choose the VAR model of lag 1 for its general character and its proved power to reveal patterns in finance [16,23]. More complex models might have higher power to reveal nuance patterns, but at the expense of a loss of generality owing to the focus on particular systems.

We include all the time series in a single model to avoid the false positives associated with pairwise Granger tests. To ensure the correct application of the VAR model, we need to verify that our analysis is consistent with its fundamental assumptions: (i) that the elements of $V(t)$ do not have a unit root, and (ii) that the error term ϵ has no temporal nor structural correlations. We verify the first set of assumptions on the properties of $V(t)$ by applying a set of tests and transformations prior to the application of the VAR model. We ensure that our conclusions are robust to the second set of assumptions by correcting for correlations in the noise term, as explained in the Material and methods section.

2.2. Impulse analysis

The VAR weights ϕ are only informative when there are no correlations in the error term ϵ of equation (2.2), which is usually not the case in practice. To extract stylized facts that can be used in the design of trading strategies, we perform an impulse analysis by measuring impulse response functions (IRF) [38] while correcting for correlations in the empirical error. This method simulates the system dynamics when it receives a shock in one of the variables, applying the VAR dynamics of equation (2.2) to reproduce the changes in the rest of the variables through time. By recording the changes in each variable, we can estimate the total size and the timespan of the perturbation produced by the shock. In essence, the IRF method creates a computational equivalent of the system under scrutiny, to test its reaction to exogenous impulses in each of its elements.

2.3. Trading strategy design and evaluation

The output of the impulse analysis step, shown in figure 1, is a set of patterns of Granger-type 'causation', i.e. it tests the null hypothesis of the absence of temporal correlations among the variables. We use these patterns as stylized facts that indicate which variables precede changes in price returns. For example, if variable $Y(t)$ has a significant impact on $\mathrm{Ret}(t)$ in the impulse analysis, we will include $Y(t)$ in our trading strategy design with sign s_Y, which takes the value 1 if the response of $\mathrm{Ret}(t)$ to $Y(t)$ was positive, and -1 otherwise. Thus, a predictor based on $Y(t)$ would be

$$\mathrm{sign}(\mathrm{Ret}(t+1)) = \mathrm{sign}(s_Y \times (Y(t) - Y(t-1))). \tag{2.3}$$

This way, we predict increases (decreases) in price between time t and $t+1$ if signals with positive responses increase (decrease) between time $t-1$ and t, and vice versa for signals with negative responses.

Downloaded from http://rsos.royalsocietypublishing.org/ on September 15, 2016

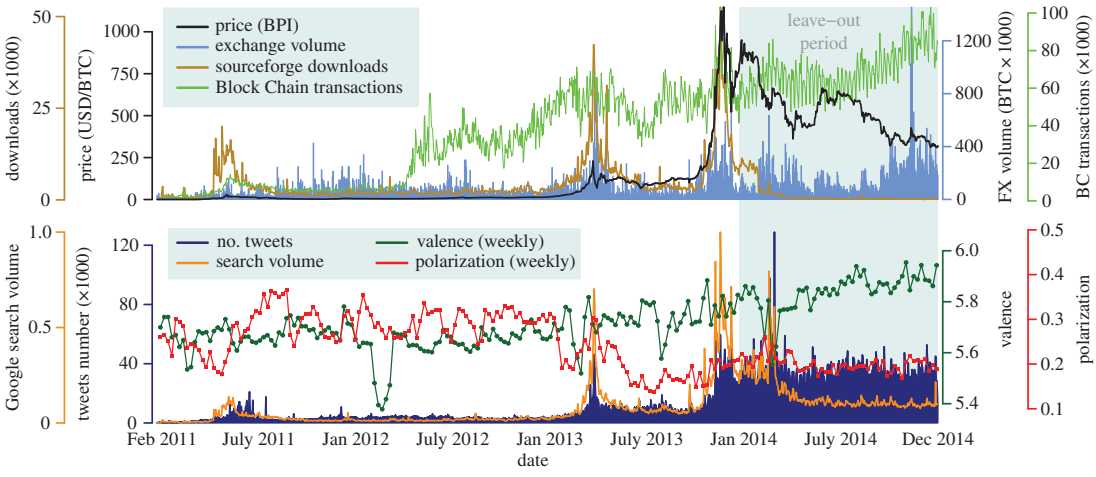

Figure 2. Time series of data volumes in the Bitcoin ecosystem. Interactive version: www.sg.ethz.ch/btc.

Since our multidimensional analysis is robust to confounds between multiple time series, the findings of impulse analysis can be integrated in a *Combined* strategy based on a voting mechanism. The *Combined* strategy applies the other predictors and formulates a prediction corresponding to the sign of the sum of their outputs, i.e. the majority vote.

We evaluate the profitability of the designed strategies in comparison to the benchmark of standard strategies, based on the backtesting over the leave-out sample as indicated in figure 1. For each strategy, we make a data-driven simulation of a trader following that strategy, and we record the profits of that trader on a daily basis. Details on the computational simulation of financial traders can be found in the Material and methods section.

2.4. Bitcoin social economic and signals

We apply our approach to the case of trading Bitcoin based on social and economic signals of the Bitcoin ecosystem. We set up a system that monitors different data sources, retrieving data in real time in combination with historical time series. The data volumes recorded during our study period of almost 4 years are shown in figure 2, and can be interactively browsed in our online visualization (www.sg.ethz.ch/btc). The signals we measure, explained more in detail in the Material and methods section include economic signals of price $P(t)$ and returns $\mathrm{Ret}(t)$, trading volume in a wide range of Bitcoin exchange markets $\mathrm{FX}_{\mathrm{Vol}}(t)$. Furthermore, we measure the economic signal of transaction volume in the Block Chain $\mathrm{BC}_{\mathrm{Tra}}(t)$, which measures the volume of usage of Bitcoin as a currency, and the amount of downloads of the most important Bitcoin client $\mathrm{Dwn}(t)$ as a measure of growth in adoption of the Bitcoin technology. The social signals we measure are the level of search volume in Google for the term 'bitcoin' $S(t)$, the word-of-mouth level as measured by the amount of *tweets* containing Bitcoin-related terms $T_N(t)$, and the emotional valence $T_{\mathrm{Val}}(t)$ and opinion polarization $T_{\mathrm{Pol}}(t)$ expressed in those tweets using lexicon-based approaches from psycholinguistics [39,40] (more details in Material and methods). All these signals are shown in figure 2, illustrating the large oscillations of price and other signals related to Bitcoin.

3. Results

3.1. Data-driven Bitcoin trading strategy design

For our statistical analysis, we include all the data up to 1 January 2014, covering almost 3 years. After applying stationarity tests, we conclude that the time series of price returns $\mathrm{Ret}(t)$ can be assumed to be stationary, as well as the first differences of the other seven signals (details on the stationarity test results can be browsed in www.sg.ethz.ch/btc and in the electronic supplementary material). As a consequence, we define our variable vector as

$$V(t) = [\mathrm{Ret}(t), \Delta\mathrm{FX}_{\mathrm{Vol}}(t), \Delta\mathrm{BC}_{\mathrm{Tra}}(t), \Delta\mathrm{Dwn}(t), \Delta S(t), \Delta T_N(t), \Delta T_{\mathrm{Val}}(t), \Delta T_{\mathrm{Pol}}(t)],$$

Downloaded from http://rsos.royalsocietypublishing.org/ on September 15, 2016

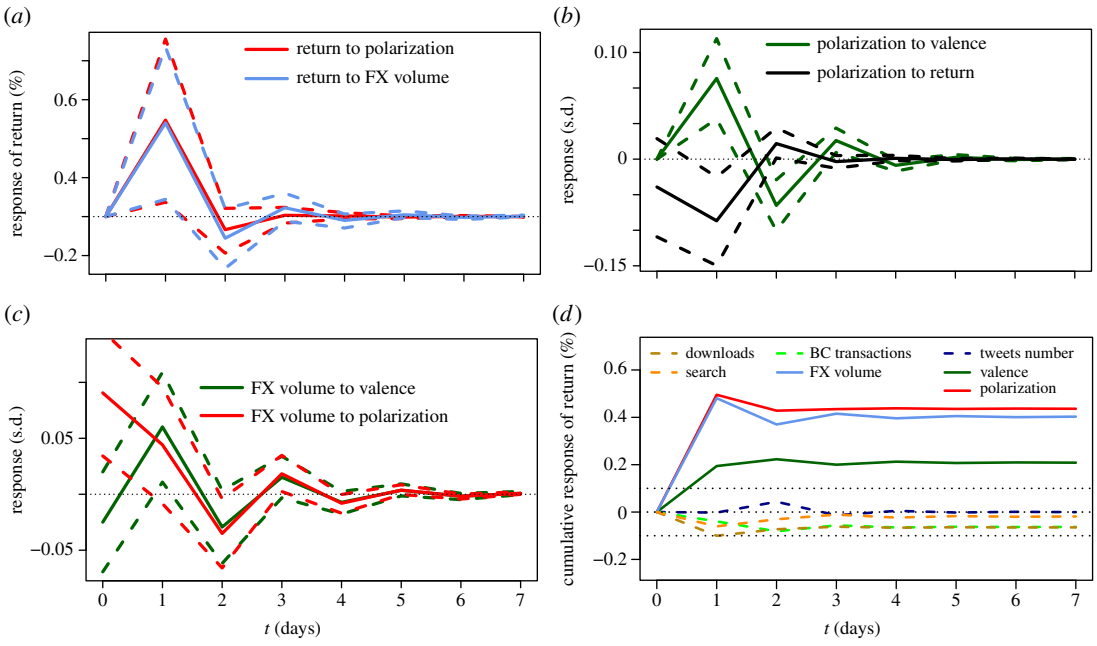

Figure 3. Results of IRF analysis. (*a*) IRF of return to shocks in Twitter polarization and exchange volume, (*b*) of Twitter polarization to shocks in return and Twitter valence, and (*c*) of exchange volume to shocks in Twitter valence and polarization (right). Solid lines show responses, dashed lines show 95% confidence intervals. (*d*) Cumulative IRF of price return to changes in the other signals. Dashed lines indicate responses below the 0.1% level.

composing the input to the multivariate analysis of our framework. We fit a VAR as explained in Material and methods over the analysis period. We compute IRF for all pairs of variables, all results including VAR estimates and IRF values can be browsed in www.sg.ethz.ch/btc and in the electronic supplementary material. Here, we comment on the most relevant results, which serve as input for our trading strategy design.

Figure 3*a* shows the IRF of returns to shocks in polarization and volume in exchange markets, where the response is measured in return percentages. Both polarization and exchange volume have significantly positive influences in price returns 1 day after the shock, decreasing rapidly afterwards. The increase of returns with polarization is consistent with the hypothesis that disagreement fuels trading in speculative scenarios [30,31], where information asymmetries fuel price bubbles. Exchange volume also increases with polarization, as shown in figure 3*c*, but the relationship is instantaneous rather than lagged as in the case of returns.

Figure 3*b* shows the response of polarization in Twitter to shocks in returns and valence. The negative effect of polarization shows that price drops lead to increases in polarization, signalling the disagreement in the Bitcoin community owing to price crashes. The pattern linking valence to polarization is relevant, revealing that periods with increasing positivity in expression precede stages of higher polarization. The role of valence can further be observed in the IRF of exchange volumes in figure 3*c*, in which valence has a significant effect. The combination of patterns of increasing polarization and exchange volume following stages of increasing valence show the relevance of valence in price returns, in addition to the effects of polarization and exchange volume.

We further validated these results in two ways. First, we fit a VAR with lags longer than a day, selecting the optimal lag that optimizes the Bayesian information criterion. We found that a lag of 2 is optimal, but the results of the fits and IRF analysis did not qualitatively change (see the electronic supplementary material). Second, we performed a Monte Carlo test, computing the IRF for time series with randomized permutations of the values. The results of these permutation tests are consistent with the above results, as reported in the electronic supplementary material, showing the robustness of our approach.

3.2. Turning analysis into strategy

We summarize the above findings as stylized facts that can drive the decisions of an algorithmic trader. We focus closer on the role of each signal into returns, by computing the cumulative changes given by

Downloaded from http://rsos.royalsocietypublishing.org/ on September 15, 2016

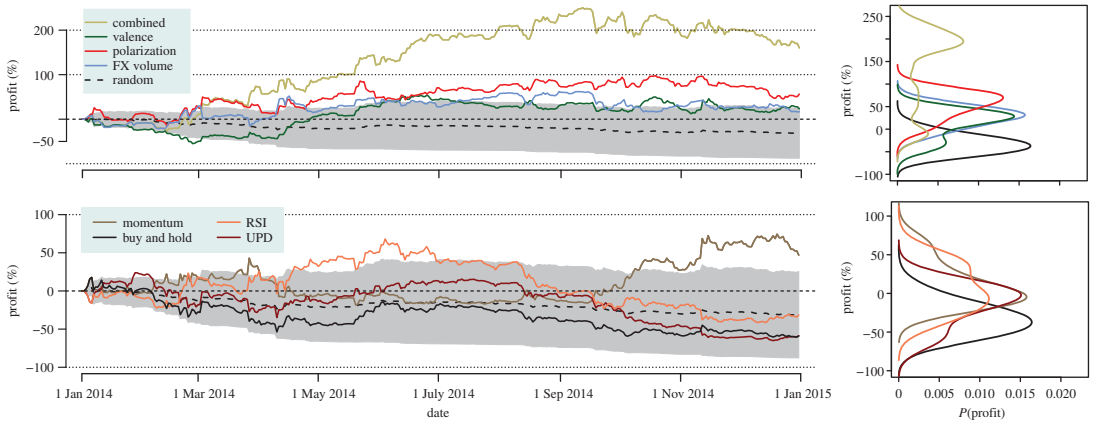

Figure 4. Profits of trading strategies. Left: time series of profit for our strategies (top) and technical strategies (bottom). Shaded areas show 1 s.d. of the random strategy. Interactive version: www.sg.ethz.ch/btc. Right: kernel density plots of the profit of each strategy (bandwidth = 15%).

the IRF analysis. This way, we can identify which signals show a sizable pattern that precedes changes in returns, and filter out those that are not significant or can be explained as confounds of the others. Figure 3*d* shows the results, measuring the cumulative change in return percentage when each one of the other signals receives a shock of size 1 s.d. The three signals with effects above the 0.1% level are polarization, valence and exchange volume, reaching effects up to 0.5% in 1 day that prevail through time. Note that this is a relatively large value, because trading results in multiplicative returns. Such effect sizes have strong potential impact on the profitability of trading strategies over long time periods. This allows us to discard the rest of the signals, feeding into our trading strategy design by producing four strategies: three strategies of positive sign, *Valence*, *Polarization* and *FXVolume*, and a fourth *Combined* strategy determined by a voting mechanism as explained in the Trading strategy framework section.

3.3. Bitcoin strategy evaluation

To evaluate the profitability of our four strategies, we set up a benchmark against random strategies and technical strategies, using the actual exchange rate of BTC for USD in `bitfinex.com` as well as the Bitcoin Price Index (BPI; see www.sg.ethz.ch/btc for results with BPI). Random strategies sample a random number with 0 mean at every time t and formulate a prediction based on the sign of the random number. Among technical strategies, the simplest is *Buy and Hold*, which simply buys BTC with the initial capital at time $t = 1$, selling it only once at the time when profits are evaluated. The technical strategies we use are a benchmark of simple standard predictions [5]: (i) the *Momentum* strategy, which predicts that price changes at time $t + 1$ will be the same as at time t, (ii) the up and down persistency strategy *UPD*, which predicts that price increases at time t are followed by decreases at time $t + 1$, and vice versa, and (iii) the relative strength index strategy *RSI*, which computes an additional time series of ratios of return sign frequencies over a rolling window of 5 days, and predicts price changes based on reversals of this time series (more details in [5]).

The simulation of each strategy produces a time series of profits:

$$\text{Profit}(t) = \frac{C(t) - C(0)}{C(0)} \times 100, \tag{3.1}$$

where $C(t)$ is the capital of the trader at time t and $C(0)$ is the initial investment capital. Figure 4 shows the time series of profits for our four strategies and the technical strategies. In addition, we compute the profit of *Buy and Hold*, and the results of the simulation of 10 000 random traders. The *Valence*, *Polarization* and *Combined* strategies clearly perform better than a random trader, while the *FXVolume* is not very far from the result of random traders. Among the technical strategies, only *RSI* and *Momentum* are able to eventually reach beyond the outcome of random traders, but are still clearly outperformed by the *Polarization* and *Combined* strategies.

The stopping time of the simulation of a trading strategy is given by our data, but a variety of factors might trigger a trader to stop trading earlier in a real scenario [3]. For that reason, we explore the distribution of profits of each strategy, assuming that the trading stops at any arbitrary point of

Downloaded from http://rsos.royalsocietypublishing.org/ on September 15, 2016

Table 1. Sharpe ratios and mean daily returns of strategies.

	Combined	Polarization	Valence	FXVolume	Buy and Hold
SR	1.7653	1.0120	0.6410	0.5738	−0.7741
μ_R	0.3229	0.1779	0.1183	0.1082	−0.1635
	Momentum	UPD	RSI	DJIA	random
SR	0.9146	−0.8990	−0.1772	0.7995	−1.6590
μ_R	0.1625	−0.1736	−0.0346	0.0345	−0.0963

our backtesting period. Thus, for each strategy, we have a set of profit values, one for each possible trading end date. The right panel of figure 4 shows the kernel density plots of the distributions of profits for each strategy. It can be appreciated that the most profitable strategy is *Combined*, followed by *Polarization* and then *Valence* and *RSI*. We quantitatively assessed this result, through Wilcoxon tests [41] over the distributions of profits (more details in the electronic supplementary material), confirming the observation that the most profitable strategies are *Combined*, and *Polarization*. More precisely, the *Combined* strategy gives profits beyond 100% for most of the time during the trading period.

While surveying cumulative returns is illustrative of the performance of the strategies, the multiplicative nature of cumulative returns overweights early positions and is biased towards the beginning of the evaluation period. To properly evaluate trading strategies, we calculated the Sharpe ratio [42], measuring risk-corrected profits as: $\mathrm{SR} = (\mu_R - R_f)/\sigma_R$, where μ_R and σ_R are the mean and standard deviation of the daily rate of return of a strategy $R(t) = (C(t) - C(t-1))/C(t-1)$. R_f is the 'risk-free' return rate of a theoretical investment that would give certain profit under no risk at all, which is often estimated as the interest rate of high-quality sovereign bonds. At the time of writing, some European sovereign bonds are giving interest rates close to zero or even negative [43], which motivates our conservative choice of $R_f = 0$. The value of SR is calculated in annualized units, taking into account that Bitcoin can be traded 365 days a year.

Table 1 reports the Sharpe ratio SR and the mean daily return μ_R for all strategies, as well as for the *DJIA* and the average of 10 000 random traders. The Sharpe ratio analysis is consistent with the results of the cumulative returns analysis, showing that the *Combined* strategy provides the highest returns, with the best SR value above 1.75 and with daily returns above 0.3% per day. The profitability of these strategies illustrate how social media sentiment can produce positive returns on investment, especially when including polarization measures beyond the trivial quantification of valence or mood.

3.4. Costs and risks of the Combined strategy

To better understand the possible weaknesses of the *Combined* strategy, we run a series of tests to evaluate the role of trading costs and additional risks. Trading Bitcoin in an online market usually comes at a cost, which often depends on the activity and the traded capital. These trading costs should not be confused with the transaction fees in the Block Chain [9], which do not depend on the transacted cost and are not associated to any market of exchange to other currencies. Trading costs can potentially erode the profitability of trading strategies, especially if they require many movements. We simulated the same backtests for costs increasing from 0 to 0.3% of the exchanged capital, a value well above the maximum costs of major trading platforms [44]. As a simplification, we assume that buying, selling and borrowing costs are the same, yet their values might depend on the trading volume of a strategy [44]. Figure 5 shows the final profits of the *Combined* strategy, which decrease monotonically with trading costs. The strategy is still highly profitable for low costs, but for costs above 0.25% the strategy is not profitable any more. Furthermore, we repeat this analysis assuming the limitation that daily positions need to be forcefully closed at the end of each trading period (shown in the electronic supplementary material), finding a decrease in returns but that the strategy is still profitable for trading costs of 0.1%, a typically high cost of current exchange platforms [45].

In this application, the leave-out period is characterized by decreasing BTC prices. Thus, it is important to evaluate the role of each possible trading action: longing when BTC are bought to be sold later, and shorting when they are borrowed and traded as explained above. We repeat the backtesting of the *Combined* strategy allowing only short and only long positions, following the methodology of Preis *et al*. [6]. As reported more in detail in the electronic supplementary material, the only short strategy

Downloaded from http://rsos.royalsocietypublishing.org/ on September 15, 2016

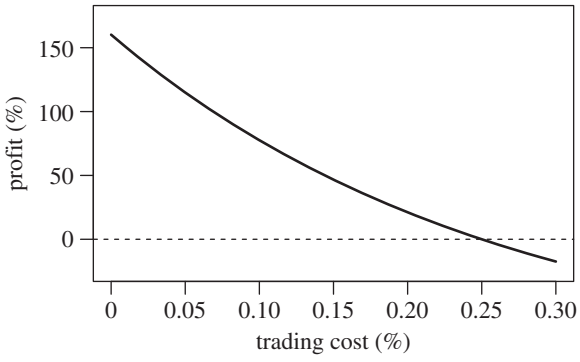

Figure 5. Final profit of the *Combined* strategy versus trading cost.

yields higher cumulative returns than the only long strategy, as expected from a period in which prices decrease steadily.

We test further properties of the behaviour of the *Combined* strategy in the leave-out period. The distribution of daily returns of the *Combined* strategy during the leave-out period follows a lognormal distribution, as tested through maximum-likelihood fits and Kolmogorov–Smirnov tests (see the electronic supplementary material). The time series of returns of this strategy is also not autocorrelated and can be considered stationary (see the electronic supplementary material for stationarity tests of daily returns). This additional analysis shows that the high profitability of the *Combined* strategy is not owing to risky correlations in the behaviour of the trading strategy.

4. Concluding remarks

Our work applies established methods of time-series analysis and computational finance to integrate the analysis, design, and evaluation of trading strategies and social and economic signals. We have shown that our approach successfully reveals temporal patterns in the Bitcoin ecosystem, in particular the relationship between price returns and the signals of exchange volume and Twitter valence and polarization. Our statistical analysis is robust to noise correlations and the finite nature of time series, providing a consistent set of results that we can apply to strategy design. We evaluated the profitability of our strategies through data-driven simulations of a computational model of a trader, showing that a strategy which combines valence, polarization and exchange volume can reach very high profits in less than a year. The added value of including polarization in our analysis constitutes evidence that collective factors of emotions and opinions have the potential to predict financial returns, beyond trivial macroscopic aggregates like average valence.

Our framework can be applied to other trading scenarios in which social signals are available, like in the case of company stock trading driven by sales data, news information and social media sentiment towards a company. The general nature of our methods are of special relevance for real trading scenarios, as the stylized facts we use to design strategies provide a tractable explanation for their mechanisms. This allows traders to understand and evaluate the principles of the algorithmic trading strategies designed in our framework. Such tractability is an advantage in comparison to more complex, nonlinear, or subsymbolic models that do not have straightforward interpretations. Nevertheless, improvements can be expected from the addition of longer time lags, higher frequency trading and real-time optimization approaches. Furthermore, the rules that drive our trading strategies do not require retraining or calibration during trading, and the social and economic signals we employ can be quantified during a day in order to have an instant trading decision ready at the beginning of the next day. Our application to Bitcoin trading is thus realistic, making use of shorting options and performing well under the typical trading costs of Bitcoin markets [44].

The application of our results should be taken with caution. Historical profit through backtesting do not necessarily predict future ones, and the information sources analysed here could be adopted by Bitcoin traders. Our evaluation goes as far as the representativity of the leave-out sample, and future research should evaluate the performance of our approach when prices rise and when traders are aware of the existence of our trading strategies. Financial markets are known to quickly absorb knowledge, as it happened with the inclusion of search trends data in stock trading [7]. It is also difficult to estimate the scalability of automatic trading strategies, as financial markets are complex adaptive systems that

Downloaded from http://rsos.royalsocietypublishing.org/ on September 15, 2016

react to trades of large volume. Furthermore, systemic risk emerges from algorithmic trading, creating *flash crashes* owing to algorithmic resonance [8]. In addition, structural changes and additional risks in borrowing and lending Bitcoin for shorting can emerge when exchange markets close or governments regulate Bitcoin, changing the rules of the game in a way such that our trading strategies might not work anymore.

With our study, we have shown that it is possible to turn social signals into profit. This extends the range of typical business applications for social media data like viral marketing or user engagement. Specifically, our combination of statistical analysis and backtesting serves as a framework for future applications of social media data in algorithmic trading. It allows a robust validation of strategy profits and a clear understanding of the system dynamics behind these profits. The application of our framework to Bitcoin trading illustrates that (asymmetric) information and profit are two manifestations of the same thing, and how traders can apply these macroscopic information sources to derive large profits. We foresee that the applications of social signals to finance will reach far beyond Bitcoin, not only to make private profit but also to understand the dynamics of individual and collective decisions and emotions.

5. Material and methods

5.1. Stationarity tests

Before fitting the VAR model, we test the stationarity of each time series through two alternative tests: (i) the augmented Dickey Fuller (ADF) test [46], which has the null hypothesis that the tested time series is *non-stationary*, and (ii) the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test [47], with the null hypothesis that the time series is *stationary*. Under these two tests, it can be considered safe that a time series is stationary if it passes the ADF test with a p-value below 0.05 and does not pass the KPSS test, giving a p-value above 0.1 [16,38]. We first analyse the time series of levels of each signal $X(t)$, applying the differentiation operator $\Delta X(t) = X(t) - X(t-1)$ until each time series is stationary. This step is inspired in the Box–Jenkins method of ARIMA time-series analysis [48], and it is usual to reach stationarity after first differences [16,38]. The stationary properties of these time series imply that their means and standard deviations are bound, allowing us to renormalize them through the Z-transformation $Z(t) = (X(t) - \mu_X)/\sigma_X$, where μ_X and σ_X are the mean and standard deviation of each time series. This way, all time series have the same scale and variance, and their effects in statistical analysis can be compared.

5.2. Impulse response function analysis

In the impulse analysis, we correct for the correlations in ϵ in two ways. First, we apply orthogonalized impulses of unit covariance, creating a shock of 1 s.d. in a variable under the error correlations of the VAR [49]. Second, we apply bootstrapping on the resulting responses by producing surrogate time series from resampling the residuals [38]. This way, we numerically compute confidence intervals of the responses in a very strict way, avoiding false positives and taking into account the finite size of the analysis period. In our case, we create 10 000 bootstrap samples to estimate 95% confidence intervals of the responses. As a result, we simultaneously measure the dynamics of the system and test their statistical significance.

5.3. Trading based on predictions

During each timestep, the prediction function makes a forecast either based on equation (2.3) or based on the price time series for technical strategies. Positive predictions translate into *buy* decisions when the trader does not own the asset, and *hold* if it does. When the predictor takes value 0, no change is done and the previous position is imitated. Negative predictions translate into *sell* positions when the trader owns the asset or *short* when it does not own it. Shorting works as follows: traders can make profit from correct predictions of price drops even if they do not own the asset predicted to drop in price. This is implemented by borrowing the asset, selling it first and buying it later for a lower price. The limitation for borrowing is usually imposed on the amount of capital already held by the trader, and often incurs in additional trading costs and legal regulations [50]. The simulation of each strategy produces a time series of profits, allowing us to measure their profitability based on historical data.

Buy and sell orders have respective costs c_b and c_s, which are proportional to the total traded capital. In our case, we assume all costs are equal $c = c_\mathrm{b} = c_\mathrm{s}$, leaving particular realizations of the costs as open

Downloaded from http://rsos.royalsocietypublishing.org/ on September 15, 2016

for future research. We compute daily cumulative returns when trading stops at $t+1$, holding USD or selling BTC at the price of $t+1$. Our trading simulations have a limit on short selling set by the amount of capital held by the trader and assume that short selling needs to be instantly executed, i.e. short positions are limited to one iteration. In summary, the strategy we execute is a single-asset backtesting scenario in which 100% of the capital is invested at each time step and shorting is limited. The pseudocode of this simulation is shown in algorithm 1.

Algorithm 1: Algorithm of trading simulation.

```
nUSD = 1; CR[1]=1;
for each t from 1 to T-1 do
    if prediction(X,t) == 1 and nBTC == 0 then
        nBTC = nUSD * (1-c_b) / P[t] ;
        nUSD = 0
    else if prediction(X,t) == -1 and nBTC > 0 then
        nUSD= nBTC * (1-c_s) * P[t] ;
        nBTC = 0
    else if prediction(X,t) == -1 and nBTC == 0 then
        nBTCb = nUSD / P[t] ;
        nUSD = nUSD + nBTCb * (1-c_s) * P[t] - nBTCb * P[t+1] / (1-c_b)
    CR[t+1] = nUSD + nBTC * P[t+1] *(1-c_s);
```

5.4. Economic signals from financial data

The establishment and bankruptcy of various Bitcoin exchange markets motivated the creation of the BPI [51]. The BPI combines a set of price indices from well-performing exchange marketplaces to provide a reference for BTC/USD exchange rates and is accepted as a standard measure of Bitcoin price in economics [18,52,53]. We use the daily closing prices of each day t at 23.59 GMT from `coindesk.com`, composing the time series of price $P(t)$ from 1 February 2011 to 31 December 2014, shown in the top panel of figure 2. The BPI is not necessarily tradeable, and for that reason we evaluate our trading strategies with the actual exchange rate of BTC for USD in `bitfinex.com`, one of the largest markets reported in `coindesk.com`. We also retrieved the daily volume of BTC exchanged in 80 online markets for other currencies from `bitcoincharts.com`. Aggregating all these data sources, we compose an Internet-wide measurement of Foreign eXchange (FX) volume of BTC traded every day $\mathrm{FX}_{\mathrm{Vol}}(t)$, including more than 152 million BTC in exchange trades as we recorded in early 2015.

Every purchase of products and services in BTC leaves a trace in the *Block Chain*, the distributed ledger that records all transactions in the Bitcoin network. We construct a time series with the daily amount of Block Chain transactions $\mathrm{BC}_{\mathrm{Tra}}(t)$, as measured by `blockchain.info` every day at 18.15.05 UTC, which we approximate to 00.00 GMT of the next day. While some data are lost in this additional delay of a few hours, further research can provide more precise measurements up to the minute level using the raw information in the *Block Chain* itself as in [16]. This way, we include more than 55 million transactions in the studied period, measuring the overall activity of the system when using Bitcoin as means of exchange. In addition, we measure the growth of the Bitcoin market through its amount of adopters, using the operationalization of measuring the amount of downloads of the most popular Bitcoin client (http://sourceforge.net/projects/bitcoin) [16], daily binned in line with other time series. The resulting time series of downloads $\mathrm{Dwn}(t)$ is shown in the top panel of figure 2.

5.5. Social signals

We record the overall interest towards Bitcoin through information search, as quantified by the Google trends volume for the term 'bitcoin', $S(t)$, as recorded in early 2015 and shown in the bottom panel of figure 2. We choose the search term 'bitcoin' instead of the 'Bitcon - Currency' topic, which was introduced as a functionality of Google Trends during our analysis period. While the topic approach can be more precise for demographics and motive analysis [12], we follow a homogeneous approach including only the term trend data that was available during the whole study period. It is important to note that Google Trends data is provided with an additional lag of 1 day and on the basis of Pacific Standard Time instead of GMT, adding almost another day of lag. While this is not an issue for the

Downloaded from http://rsos.royalsocietypublishing.org/ on September 15, 2016

historical analysis, the evaluation of any trading strategy using $S(t)$ needs to take into account this additional delay.

We track the attention of social media about Bitcoin in Twitter via the Topsy data service (http://topsy.com/). From the full track of data accessible by Topsy [54], we focus on tweets that contain Bitcoin terms as in previous research [16], finding a total of 19 578 671 Bitcoin-related tweets. The first social signal we extract from Twitter is the daily amount of unique tweets about Bitcoin $T_N(t)$ binned in 24 h windows starting at 00.00 GMT, measuring the level of word-of-mouth and attention towards Bitcoin and shown in the bottom panel of figure 2. We continue by measuring the collective emotional valence with respect to Bitcoin, as expressed through the text of Bitcoin-related tweets. Valence is considered the most important dimension of affect, quantifying the degree of pleasure or displeasure of an emotional experience [55]. The expression of valence through text is a common practice in psychological research, in which lexicon techniques are used to empirically measure emotions [56,57]. We measure the average daily valence of Bitcoin-related tweets through a state-of-the-art lexicon technique [40], which improves the previous ANEW lexicon method [56] with more than 13 000 valence-coded words. We compute the daily average Twitter valence about Bitcoin during day t in two steps: First, we measure the frequency of each term in the lexicon during that day, and second, we compute the average valence weighting each word by its frequency. This measurement matches more than 50 million valence-carrying tokens and produces the time series of Twitter valence $T_{\mathrm{Val}}(t)$.

Our last social signal is opinion polarization, which builds up on measuring the semantic orientation of words into positive and negative evaluation terms [58]. We apply the LIWC psycholinguistics lexicon-based method [39] and expand its lexicon of stems into words by matching them against the most frequent English words of the Google Books dataset [59]. As a result, we consider 3463 positive and 4061 negative terms that appear as more than 8 million Twitter tokens. We compute the daily polarization of opinions in Twitter around the Bitcoin topic $T_{\mathrm{Pol}}(t)$, calculating the geometric mean of the daily ratios of positive and negative words per Bitcoin-related tweet. Note that, instead of repeating a measurement of valence through two different lexica, we quantify polarization as a complementary dimension to emotional valence. This way, opinion polarization measures the simultaneous coexistence of positive and negative subjective content, rather than its overall orientation [23,58].

Ethics. This research is based on observational data shared publicly. No personal or individual information was retrieved, stored or analysed.

Data accessibility. All data used for this article is publicly available through the corresponding Application Programming Interfaces. Processed data and scripts to reproduce these results are available at https://www.sg.ethz.ch/btc.

Authors' contributions. D.G. conceived the research, retrieved data and performed the analyses. D.G. and F.S. wrote the manuscript.

Competing interests. We declare we have no competing interests.

Funding. This work was funded by the Swiss National Science Foundation (CR21I1_146499).

Acknowledgements. We thank Emre Sarigol for his technical assistance.

References

1. Lazer D *et al.* 2009 Life in the network: the coming age of computational social science. *Science* **323**, 721–723. (doi:10.1126/science.1167742)
2. Schweitzer F, Vespignani A. 2012 Editorial. *EPJ Data Sci.* **1**, 1. (doi:10.1140/epjds1)
3. Treleaven P, Galas M, Lalchand V. 2013 Algorithmic trading review. *Commun. ACM* **56**, 76–85. (doi:10.1145/2500117)
4. Park C-H, Irwin SH. 2004 The profitability of technical analysis: a review. Technical report, University of Illinois at Urbana-Champaign, Department of Agricultural and Consumer Economics, Champaign, IL, USA.
5. Biondo AE, Pluchino A, Rapisarda A, Helbing D. 2013 Are random trading strategies more successful than technical ones? *PLoS ONE* **8**, e68344. (doi:10.1371/journal.pone.0068344)
6. Preis T, Moat HS, Stanley HE. 2013 Quantifying trading behavior in financial markets using *Google Trends*. *Sci. Rep.* **3**, 1684. (doi:10.1038/srep01684)
7. Curme C, Preis T, Stanley HE, Moat HS. 2014 Quantifying the semantics of search behavior before stock market moves. *Proc. Natl Acad. Sci. USA* **111**, 11 600–11 605. (doi:10.1073/pnas.1324054111)
8. Cusumano MA. 2014 The Bitcoin ecosystem. *Commun. ACM* **57**, 22–24. (doi:10.1145/2661047)
9. Nakamoto S. 2008 Bitcoin: a peer-to-peer electronic cash system. Technical report, Bitcoin Foundation. See https://bitcom.org/bitcom.pdf.
10. Van Alstyne M. 2014 Why Bitcoin has value. *Commun. ACM* **57**, 30–32. (doi:10.1145/2594288)
11. Christin N. 2013 Traveling the silk road: a measurement analysis of a large anonymous online marketplace. In *Proc. of the 22nd Int. Conf. on World Wide Web, Rio de Janerio, Brazil, 13th–17th May 2013*, pp. 213–224.
12. Yelowitz A, Wilson M. 2015 Characteristics of Bitcoin users: an analysis of Google search data. *Appl. Econ. Lett.* **22**, 1030–1036. (doi:10.1080/13504851.2014.995359)
13. Bitpay. Over 44 000 businesses and charities accept bitcoin with bitpay. See https://bitpay.com/directory#.
14. Dell. Dell now accepts Bitcoin. See http://dell.to/1uk8ocZ.
15. Schweitzer F, Fagiolo G, Sornette D, Redondo FV, Vespignani A, White DR. 2009 Economic networks: the new challenges. *Science* **325**, 422–425. (doi:10.1126/science.1173644)
16. Garcia D, Tessone CJ, Mavrodiev P, Perony N. 2014 The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy. *J. R. Soc. Interface* **11**, 20140623. (doi:10.1098/rsif.2014.0623)
17. Kristoufek L. 2013 *BitCoin* meets *Google Trends* and *Wikipedia*: quantifying the relationship between phenomena of the Internet era. *Sci. Rep.* **3**, 3415. (doi:10.1038/srep03415)
18. Kristoufek L. 2015 What are the main drivers of the Bitcoin price? Evidence from wavelet coherence

Downloaded from http://rsos.royalsocietypublishing.org/ on September 15, 2016

analysis. *PLoS ONE* **10**, e0123923. (doi:10.1371/journal.pone.0123923)

19. Schoen H, Gayo-Avello D, Metaxas PT, Mustafaraj E, Strohmaier M, Gloor P. 2013 The power of prediction with social media. *Internet Res.* **23**, 528–543. (doi:10.1108/IntR-06-2013-0115)
20. Bordino I, Battiston S, Caldarelli G, Cristelli M, Ukkonen A, Weber I. 2012 Web search queries can predict stock market volumes. *PLoS ONE* **7**, e40014. (doi:10.1371/journal.pone.0040014)
21. Mao Y, Wei W, Wang B, Liu B. 2012 Correlating s&p 500 stocks with twitter data. In *Proc. of the First ACM Int. Workshop on Hot Topics on Interdisciplinary Social Networks Research, Beijing, China, 12–16 August 2012*, pp. 69–72. ACM.
22. Gonzalez-Bailon S, Banchs RE, Kaltenbrunner A. 2010 Emotional reactions and the pulse of public opinion: measuring the impact of political events on the sentiment of online discussions. (http://arxiv.org/abs/1009.4019)
23. Tumarkin R, Whitelaw RF. 2001 News or noise? Internet postings and stock prices. *Financ. Anal. J.* **57**, 41–51. (doi:10.2469/faj.v57.n3.2449)
24. Saavedra S, Duch J, Uzzi B. 2011 Tracking traders' understanding of the market using e-communication data. *PLoS ONE* **6**, e26705. (doi:10.1371/journal.pone.0026705)
25. Gilbert E, Karahalios K. 2010 Widespread worry and the stock market. In *Int. AAAI Conf. Web and Social Media, 23–26 May 2010, George Washington University, Washington, DC*, pp. 59–65.
26. Bollen J, Mao H, Zeng X. 2011 Twitter mood predicts the stock market. *J. Comput. Sci.* **2**, 1–8. (doi:10.1016/j.jocs.2010.12.007)
27. Garcia D, Abisheva A, Schweighofer S, Serdult U, Schweitzer F. In press. Network polarization in online politics participatory media. To appear in Policy and Internet. See http://bit.ly/1zGCPOv.
28. Garcia D, Tanase D. 2013 Measuring cultural dynamics through the Eurovision Song Contest. *Adv. Complex Syst.* **16**, 33. (doi:10.1142/S0219525913500379)
29. Hirshleifer J. 1977 The theory of speculation under alternative regimes of markets. *J. Finance* **32**, 975–999. (doi:10.2307/2326507)
30. Harris M, Raviv A. 1993 Differences of opinion make a horse race. *Rev. Financ. Stud.* **6**, 473–506. (doi:10.1093/rfs/5.3.473)
31. Antweiler W, Frank MZ. 2004 Is all that talk just noise? The information content of internet stock message boards. *J. Finance* **59**, 1259–1294. (doi:10.1111/j.1540-6261.2004.00662.x)
32. Tetlock PC. 2007 Giving content to investor sentiment: the role of media in the stock market. *J. Finance* **62**, 1139–1168. (doi:10.1111/j.1540-6261.2007.01232.x)
33. De Choudhury M, Sundaram H, John A, Seligmann DD. 2008 Can blog communication dynamics be correlated with stock market activity? In *Proc. of the Nineteenth ACM Conf. on Hypertext and Hypermedia, HT '08, Pittsburgh, PA, USA, 19–21 June 2008*, pp. 55–60.
34. Challet D, Bel Hadj Ayed A. In press. Predicting financial markets with Google trends and not so random keywords. *SSRN Electron. J.* (doi:10.2139/ssrn.2310621)
35. Ruiz EJ, Hristidis V, Castillo C, Gionis A, Jaimes A. 2012 Correlating financial time series with micro-blogging activity. In *Proc. of the Fifth ACM Int. Conf. on Web Search and Data Mining, WSDM '12, Seattle, Washington, USA, 8–12 February 2012*, pp. 513–522. (doi:10.1145/2124295.2124358)
36. Whittle P. 1953 The analysis of multiple stationary time series. *J. R. Stat. Soc. B* **15**, 125–139.
37. Adamic L, Brunetti C, Harris J, Kirilenko A. In press. *On the informational properties of trading networks*. SSRN eLibrary. See http://dx.doi.org/10.2139/ssrn.1361184.
38. Lütkepohl H. 2007 *New introduction to multiple time series analysis*. Berlin, Germany: Springer.
39. Pennebaker JW, Chung CK, Ireland M, Gonzales A, Booth RJ. 2007 The development and psychometric properties of liwc2007. See LIWC.net .
40. Warriner AB, Kuperman V, Brysbaert M. 2013 Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behav. Res. Methods* **45**, 1191–1207. (doi:10.3758/s13428-012-0314-x)
41. Wilcoxon F. 1945 Individual comparisons by ranking methods. *Biometrics Bull.* **1**, 80–83. (doi:10.2307/3001968)
42. Sharpe WF, Sharpe W. 1970 *Portfolio theory and capital markets*, vol. 217. New York, NY: McGraw-Hill.
43. Edwards B. 2014 European bond yields go negative. *Wall Street J.* Online, See http://www.wsj.com/articles/european-bond-yields-go-negative-1412093167.
44. Bitfinex. Features - short. See https://www.bitfinex.com/pages/features.
45. Bitfinex. Fees schedule. See https://www.bitfinex.com/pages/fees.
46. Fuller WA. 2009 *Introduction to statistical time series*, vol. 428. New York, NY: John Wiley and Sons.
47. Kwiatkowski D, Phillips PC, Schmidt P, Shin Y. 1992 Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *J. Econometrics* **54**, 159–178. (doi:10.1016/0304-4076(92)90104-Y)
48. Anderson OD. 1976 *Time series analysis and forecasting: the Box-Jenkins approach*. London, UK: Butterworths.
49. Zeileis A. 2004 Econometric computing with hc and hac covariance matrix estimators. *J. Stat. Software* **11**, 1–17.
50. Chance D, Brooks R. 2012 *Introduction to derivatives and risk management*. Cengage Learning, South-Western College Pub; 9th edn.
51. Coindesk. Coindesk launches proprietary Bitcoin price index. See http://bit.ly/11dDAAK.
52. Brito J, Shadab HB, Castillo A. In press. Bitcoin financial regulation: securities, derivatives, prediction markets, and gambling. *Columbia Sci. Technol. Law Rev*. (doi:10.2139/ssrn.2423461)
53. Shadab HB. 2014 Regulating Bitcoin and block chain derivatives. SSRN Preprint 2508707 .
54. Aitken T. 2013 Topsy: the social search engine allows you to scan every public tweet ever published on twitter. *CEO World Mag.* Online, See http://ceoworld.biz/2013/09/04/topsy-the-social-search-engine-allows-you-to-scan-every-public-tweet-ever-published-on-twitter.
55. Russell JA. 2003 Core affect and the psychological construction of emotion. *Psychol. Rev.* **110**, 145–00. (doi:10.1037/0033-295X.110.1.145)
56. Bradley MM, Lang PJ. 1999 Affective norms for English words (ANEW): instruction manual and affective ratings. Technical report, The Center for Research in Psychophysiology, University of Florida, Gainesville, FL, USA.
57. Garcia D, Schweitzer F. 2012 Modeling online collective emotions. In *Proc. of the 2012 Workshop on Data-driven User Behavioral Modelling and Mining from Social Media, DUBMMSM '12, Sheraton Maui, Hawaii, 2 November 2012*, pp. 37–38.
58. Osgood CE. 1964 Semantic differential technique in the comparative study of cultures. *Am. Anthropol.* **66**, 171–200. (doi:10.1525/aa.1964.66.3.02a00880)
59. Lin Y, Michel J-B, Aiden EL, Orwant J, Brockman W, Petrov S. 2012 Syntactic annotations for the Google books ngram corpus. In *Proc. of the ACL 2012 System Demonstrations, Jeju Island, South Korea, 8–14 July 2012*, pp. 169–174.