OPEN

Correspondence and requests for materials should be addressed to L.K. (kristouf@utia.cas. cz)

BitCoin meets *Google Trends* and *Wikipedia*: Quantifying the relationship between phenomena of the Internet era

Ladislav Kristoufek[1,2]

[1]Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Opletalova 26, 110 00, Prague, Czech Republic, EU, [2]Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodarenskou Vezi 4, 182 08, Prague, Czech Republic, EU.

Digital currencies have emerged as a new fascinating phenomenon in the financial markets. Recent events on the most popular of the digital currencies – *BitCoin* – have risen crucial questions about behavior of its exchange rates and they offer a field to study dynamics of the market which consists practically only of speculative traders with no fundamentalists as there is no fundamental value to the currency. In the paper, we connect two phenomena of the latest years – digital currencies, namely *BitCoin*, and search queries on *Google Trends* and *Wikipedia* – and study their relationship. We show that not only are the search queries and the prices connected but there also exists a pronounced asymmetry between the effect of an increased interest in the currency while being above or below its trend value.

I ntroduction of the Internet has completely changed the way real economy works. By enabling practically all Internet users to interact at once and to exchange and share information almost cost-free, more efficient decisions on the markets are possible. Even though the interconnection between digital and real economies has hit several bumps such as the DotCom Bubble of the break of the millennium, the benefits are believed to have overcome the costs.

One of the fascinating phenomena of the Internet era is an emergence of digital currencies such as BitCoin, LiteCoin, NameCoin, PPCoin, Ripple and Ven to name the most popular ones. A digital currency can be defined as an alternative currency which is exclusively electronic and thus has no physical form. It is also not issued by any specific central bank or government of a specific country and it is thus practically detached from the real economy. Note that a digital and a virtual currency are not synonymous since the virtual currencies are trading currencies in virtual worlds (most frequently in the massive multiplayer online games – MMOGs – such as World of Warcraft or Second Life). Even though the digital currencies are almost isolated from the real economies, their prices (exchange rates) have experienced quite an erratic behavior in the recent months. Specifically, the *BitCoin* currency – the most popular of the digital currencies – started the year of 2013 at levels of $13 per a *BitCoin* and rocketed to $230 on 9 April 2013 potentially creating an absurd profit of almost 1700% in less than four months. Later the same year, the price soared even higher to $395 on 9 November 2013, which accounts for a profit of approximately 2900% since the beginning of 2013.

Such behavior cannot be explained by standard economic and financial theories – e.g. future cash-flows model[1], purchasing power parity[2,3] and uncovered interest rate parity[4,5] – in a satisfactory manner. In general, currencies can be seen as standard economic goods which are priced by interaction of supply and demand on the market. These are driven by macroeconomic variables of an issuing country or institution (or entity in general) such as GDP, interest rates, inflation, unemployment, and others. As there are no macroeconomic fundamentals for the digital currencies, the supply function is either fixed (if the currency amount is fixed) or it evolves according to some publicly known algorithm, which is the case of the *BitCoin* market. The demand side of the market is not driven by an expected macroeconomic development of the underlying economy (as there is none) but it is driven only by expected profits of holding the currency and selling it later (as there are no profits from simply holding the currency due to no interest rates of the digital currencies). The market is thus dominated by short-term investors, trend chasers, noise traders and speculators. The fundamentalist segment of the market is completely missing due to the fact that there are no fundamentals allowing for setting of a "fair" price. The digital currency price is thus driven solely by the investors' faith in the perpetual growth. Investors' sentiment then becomes a crucial variable.
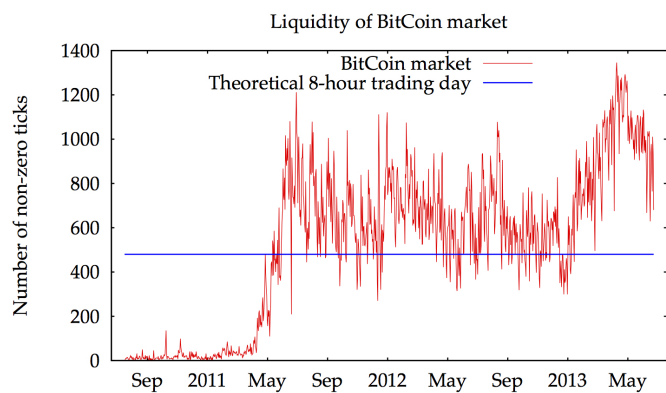
Figure 1 | Evolution of ticks number. Number of ticks with a non-zero return per day is shown. The red line represents a number of ticks for an 8-hour trading day and is shown just for illustration. It is visible that for the starting days of existence of the *BitCoin* market, there was practically no liquidity. Approximately since May 2011, liquidity has reached satisfactory levels.

However, it is not a trivial task to find a good measure or proxy of investors' sentiment in this matter. Quite recently, search queries provided by *Google Trends* and *Wikipedia* have proved to be a useful source of information in financial applications ranging from the home bias and the traded volume explanations through the earnings announcements to the portfolio diversification and trading strategies[6–12]. The frequency of searches of terms related to the digital currency can be a good measure of interest in the currency and it can have a good explanatory power.

Here, we study the relationship between prices of the *BitCoin* currency (for a detailed description of a functioning of the currency, refer to Ref. 13) and related searched terms on *Google Trends* and *Wikipedia*. We find a striking positive correlation between a price level of *BitCoin* and the searched terms as well as a dynamic relationship which is bidirectional. Moreover, we uncover an asymmetry between effects of search queries related to prices above and below a short-term trend.

Results

Dataset. We analyze the dynamic properties of the *BitCoin* currency (as the most popular of the digital currencies) and the search queries on *Google Trends* and *Wikipedia* as proxies of investors' interest and attention. Time series for the *BitCoin* currency at the most liquid market (Mt. Gox) are available since 17.7.2010 with the highest

reported frequency (a tick) of 1 minute. However, the market remained highly illiquid for approximately the first year of its existence. To separate the period into the illiquid and the liquid one, we investigate a number of ticks with a non-zero return during a specific day. Fig. 1 depicts the evolution of the *BitCoin* liquidity. As a benchmark, we also show a number of 1-minute ticks associated with an 8-hour trading day. Even though the *BitCoin* market is a 24/7 market, we use the 8-hour trading day as a simple benchmark of a liquid market. We observe that the number of ticks gets closer to the threshold value approximately in the middle of 2011. Closer inspection uncovers that since the beginning of May 2011, the number of ticks has fluctuated around the 8-hour benchmark. Therefore, we analyze the series starting on 1 May 2011 with an ending date of 30 June 2013. For *Google Trends*, we are working with weekly data and as such, we obtain 113 observations in total; while for *Wikipedia*, daily data are available so that we have 788 observations.

Evolution of both pairs – *Google Trends* (weekly) and *Wikipedia* (daily) with corresponding *BitCoin* prices – is illustrated in Fig. 2. Obviously, the daily series of *Wikipedia* entries provides a more detailed picture of the behavior of the Internet users' interest and attention together with a higher potential for a more precise statistical analysis. We observe that the prices of the digital currency are strongly correlated with the search queries of both engines. Specifically, the correlations reach the levels of 0.8786 (with $t(111) = 19.3850[<0.01]$, *p*-value is shown in the square brackets) and 0.8271 (with $t(786) = 41.2587[<0.01]$) for *Google Trends* and *Wikipedia*, respectively. The strength of these relationships is nicely illustrated in Fig. 3 where a strong linear correlation between logarithmic prices and logarithmic search frequencies is evident. The fact that such correlation is most apparent for the log-log specification is the first hint for an analysis of the logarithmic transforms rather than the original series. Moreover, the log-log specification also allows for an easy interpretation of the relationship as the elasticity. Such notion is more stressed in the next section where the stationarity and cointegration of the series are discussed.

Stationarity & cointegration. To cover various combinations of relationships, we initially study all standard transformations of the original series, i.e. the logarithmic transformation, the first differences, and the first logarithmic differences. For each of the series, we test their stationarity using the KPSS[14] and ADF[15] tests. As both tests have opposite null and alternative hypotheses, they form an ideal pair for the stationarity vs. unit-root testing. In Tab. 1, all these results are summarized. For the *BitCoin* prices (both daily and weekly), we find both the original and the logarithmic series to be
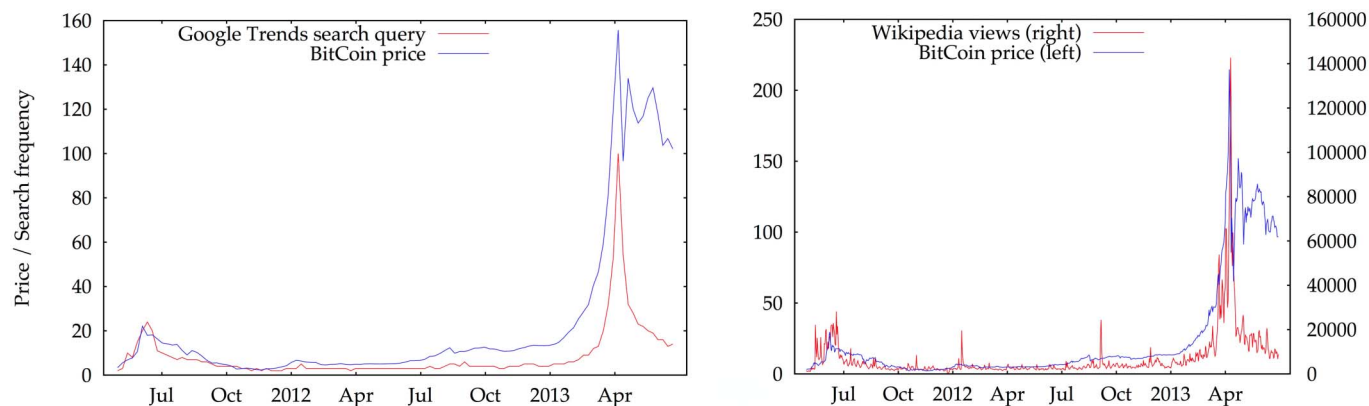

Figure 2 | BitCoin price and search queries evolution. Weekly series for *BitCoin* and *Google Trends* are shown on the left and daily series for *BitCoin* and *Wikipedia* are shown on the right. Search terms are evidently positively correlated with the prices with correlation of 0.8786 and 0.8271 for *Google Trends* and *Wikipedia*, respectively (for a log-log scale). The *BitCoin* bubble of 2013 is accompanied with rocketing search queries in both databases.
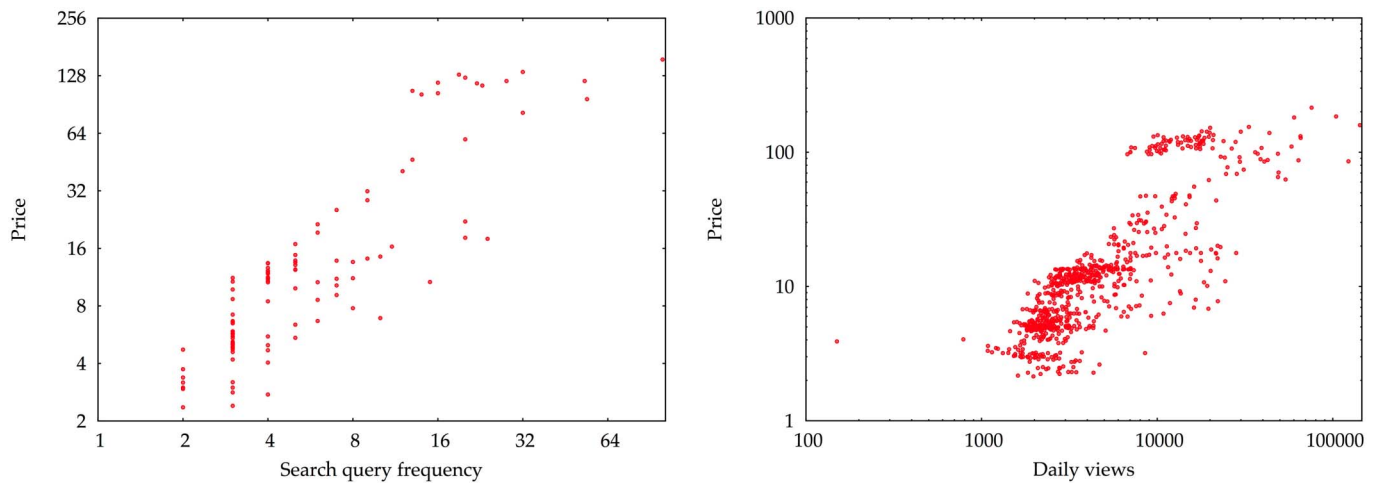
Figure 3 | Relationship between *BitCoin* price and search queries. Double logarithmic illustration of correlation between *BitCoin* prices and the searched term (*Google Trend* on the left and *Wikipedia* on the right) is shown. A positive dependence is evident and it holds for practically the whole range with correlation of 0.8786 and 0.8271 for *Google Trends* and *Wikipedia*, respectively.

non-stationary and to contain the unit-root. Correspondingly, their first differences are stationary. The same results are found for the *Wikipedia* daily views but for the *Google Trends* queries, we find the unit-root only for the logarithmic transformation of the searched terms series. For this reason and also for more convenient interpretation, we opt for the logarithmic series.

Turning now to the analysis of the dynamic properties and interconnections between the series, we are firstly interested in a potential cointegration relationship. Cointegration methodology has proved very useful in various economic and financial studies ranging from economic development[16,17] over monetary economics[18,19], international economics[20–22] to energy economics[23,24] as it enables to study a long-term relationship between series as well as their short-term dependence via the error-correction models (see the Methods section for more details). To test for the cointegration relationships, we utilize two tests of Johansen[25] – the trace and the likelihood tests. In Tab. 2, we show the results for both pairs and we find that the *BitCoin* series are not cointegrated with the *Google Trends* series but the connection to the *Wikipedia* series can be described as the cointegration. Therefore, for the first pair, we need to turn to the vector autoregression (VAR) methodology applied on the first logarithmic differences (see the Methods section for more details), and for the second pair, we stick to the standard cointegration and vector error-correction model (VECM) framework.

General results. Starting with the *Google Trends* results, we are firstly interested in the dynamic relationship between the search queries on *Google* – namely "BitCoin" (note that the search query frequency is not case sensitive so that the various versions of the word, such as "BitCoin", "Bitcoin" and "bitcoin", are included) – and price of the currency. Based on the Akaike, Hannan-Quinn and Schwarz-Bayesian information criteria, we use a single lag in the VAR approach, i.e. VAR(1) is applied on the first logarithmic differences. To control for potential autocorrelation and heteroskedasticity inefficiencies, we opt for heteroskedasticity and autocorrelation robust (HAC) standard errors. The results are summarized in Fig. 4. The charts show the response of a corresponding variable to a shock in the impulse variable. As we are working with logarithmic differences, we can interpret these shocks as a proportional reaction to a 1% shock. A 10% shock in the search queries yields a reaction of approximately 0.8% in the first and 1.2% in the second period, i.e. a total 2% reaction, and the effect vanishes for the latter periods. However, the influence also works from the opposite side and it again lasts (remains statistically significant) for two periods. The reaction to a 10% shock in search queries is followed by a total reaction of 0.8% (0.55% and 0.25% for the periods, respectively) of the prices. Putting these two together, we find that the increased interest in the *BitCoin* currency measured by the searched terms increases its price. As the interest in the currency increases, the demand increases as well causing the prices to increase. However, as the price of *BitCoin* increases so does also the interest of not only investors but also a general public. Note that it is quite easy to invest into *BitCoin* as the currency does not need to be traded in large bundles. This evidently forms a potential for a bubble development.

Turning now to the results of the *Wikipedia* daily views, we are interested in the same relationship as in the previous case but now based on the vector error-correction model (VECM) with seven lags (VECM(7)) based on the information criteria. In Fig. 5, we present the response functions which are, however, different from the previous ones as these represent permanent shifts in the response variable compared to the immediate shifts in Fig. 4. In the first 7 days (a trading week), an increase in prices causes an increasing positive reaction of the daily views. After the first week, the effect stabilizes but the interest in *BitCoin* measured by the daily views does not return back to the initial level. The complete transmission is around 0.05, i.e. a 10% change in prices is connected to a 0.5% permanent shift in the *Wikipedia* views. From the opposite side, we do not observe any statistically significant effect coming from the daily views to prices. The difference between *Wikipedia* and *Google*

Table 1	Stationarity and unit-root tests			
	KPSS	*p*-value	ADF	*p*-value
BitCoin price (daily)	5.2057	< 0.01	−0.3688	> 0.1
– log-price	6.6673	< 0.01	−0.3087	> 0.1
– difference	0.0851	> 0.1	−5.3665	< 0.01
– log-difference	0.1925	> 0.1	−4.6020	< 0.01
BitCoin price (weekly)	1.1203	< 0.01	−1.5897	> 0.1
– log-price	1.3995	< 0.01	0.3799	> 0.1
– difference	0.2127	> 0.1	−3.0070	0.0343
– log-difference	0.1694	> 0.1	−3.0938	0.0270
Google Trends frequency	0.5534	0.0370	−3.5270	< 0.01
– log	0.6942	0.0160	−1.7900	> 0.1
– difference	0.0303	> 0.1	−8.3641	< 0.01
– log-difference	0.0659	> 0.1	−3.6030	< 0.01
Wikipedia frequency	5.2057	< 0.01	−0.3688	> 0.1
– log	6.6673	< 0.01	−0.3087	> 0.1
– difference	0.0851	> 0.1	−5.3665	< 0.01
– log-difference	0.1925	> 0.1	−4.6020	< 0.01

Table 2 | Cointegration tests between *BitCoin* prices and search queries

Series	Cointegration vectors	Trace test	*p*-value	Likelihood test	*p*-value
Google Trends	**0**	**13.5640**	>**0.1**	**12.9810**	**0.0778**
	1	0.5823	> 0.1	0.5823	> 0.1
Wikipedia	0	21.6620	< 0.01	20.9150	< 0.01
	1	**0.7473**	>**0.1**	**0.7473**	>**0.1**
Wikipedia	0	125.0100	< 0.01	91.3800	< 0.01
+	1	33.6320	< 0.01	33.2540	< 0.01
feedback	**2**	**0.3782**	>**0.1**	**0.3782**	>**0.1**

Trends might be caused by the fact that of course the two engines are different and individuals using these two can have different motives and can be interested in different specifics. Nonetheless, we believe that both engines provide interesting insights into the functioning and relationship between the digital currency and a general interest in the currency. Apart from the standard effects, we are also interested whether the reaction of prices to the searched terms is symmetric, i.e. whether an increasing interest coming in hand with the increasing prices (possibly a bubble forming) has a same effect as an increasing interest connected to the decreasing prices (possibly a bubble burst).

Positive and negative feedback. A crucial disadvantage of measuring interest using the search queries on *Google Trends* or daily views on *Wikipedia* is the fact that it is hard to distinguish between interest due

to the positive or negative events. Specifically for the *BitCoin*, there is a big difference between searching for the information during an increasing trend or after the bubble burst. To separate these effects, we introduce a dummy variable equal to one if the price of *BitCoin* is above its trend level (measured by a moving average of 4 for *Google Trends* and of 7 for *Wikipedia* due to different sampling frequency) and zero otherwise. This way, we try to distinguish between a positive feedback defined as a reaction to an increasing interest (measured by search queries) while the price is above its trend value and a negative feedback defined reversely.

For the *Google Trends* pair, the results are again illustrated in Fig. 4. Here, we can see that practically the whole reaction comes from the positive feedback as there is practically no statistically significant reaction to the negative movements of the prices in a sense of the search queries. Much more interesting results are found for the
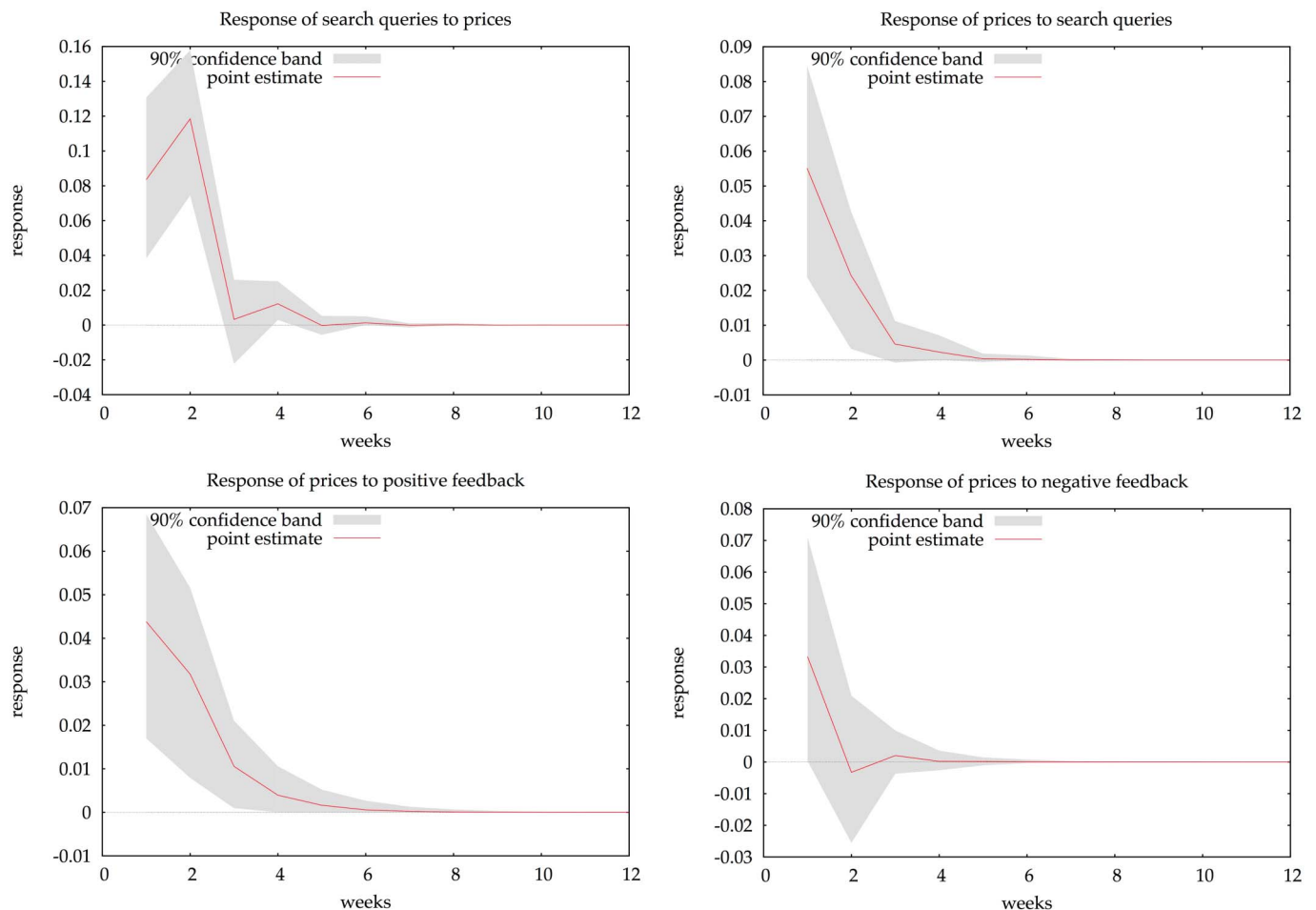

Figure 4 | **Response dynamics for *Google Trends*.** Impulse-response functions for the first logarithmic differences of *BitCoin* prices and *Google Trends* search queries. Positive relationship is evident in both directions. Responses are also partly asymmetric.
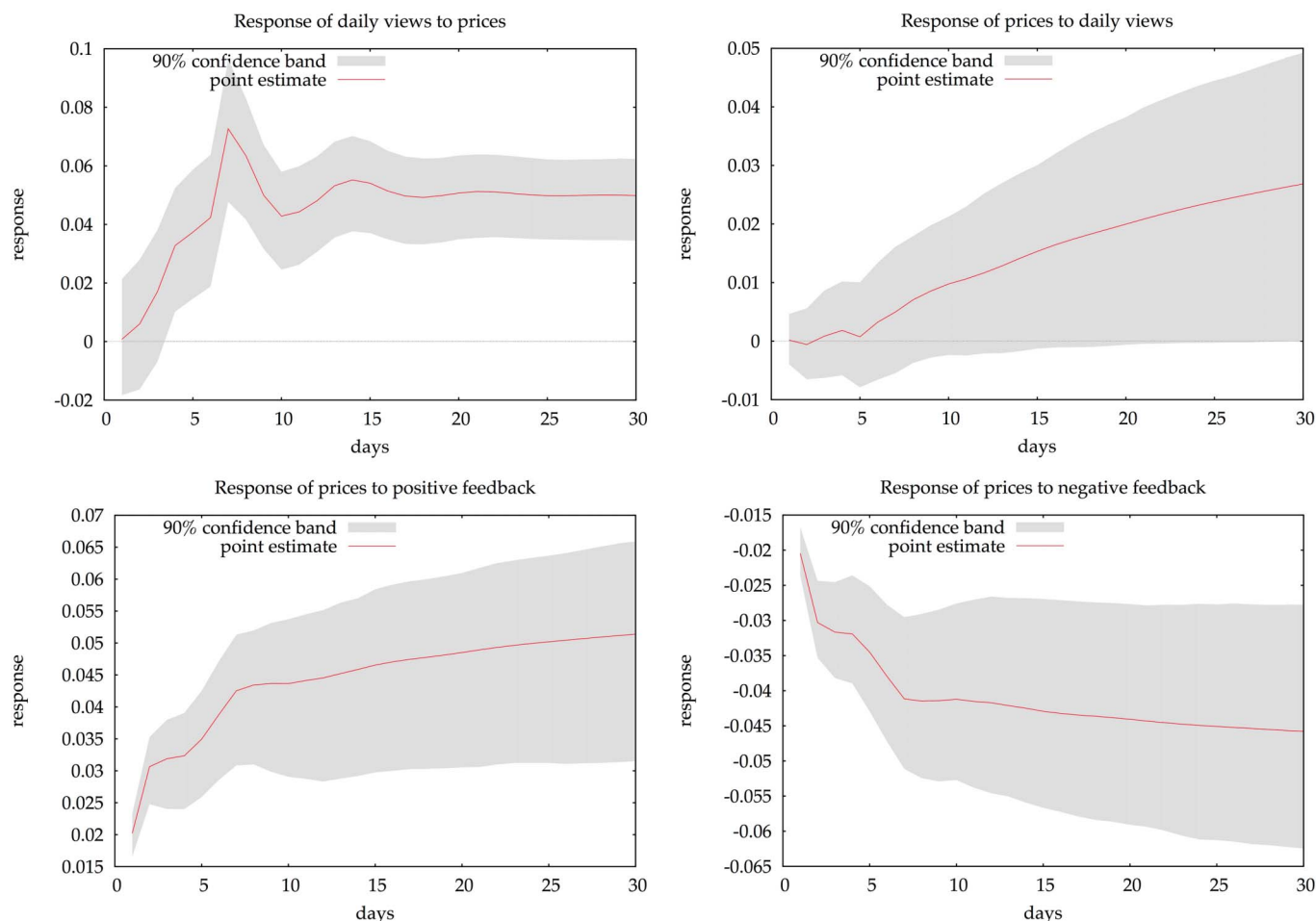
Figure 5 | Response dynamics for *Wikipedia*. Impulse-response functions for the logarithmic transformations of *BitCoin* prices and *Wikipedia* daily views. There is a positive effect of price changes on daily views on *Wikipedia* site. The opposite effect is not statistically significant. However, when the effects are separated into a positive and a negative feedback, the effect becomes statistically significant.

Wikipedia daily views. In Fig. 5, we find that the positive and negative feedback are practically symmetric around the zero reaction. That is – the reaction of prices to changes in the *Wikipedia* interest is similar for the prices being both above and below the trend but for the sign of the reaction. The complete transmission is around 0.05 and −0.05 for the positive and negative feedback, respectively. This is a crucial result because without the separation between the positive and negative feedback, we do not find any reaction of the *BitCoin* prices to the *Wikipedia* views. However, if the effect is separated, the reaction is statistically significant and of an expected sign. If the prices are going up and the public interest in the matter is growing, the prices will likely continue soaring up. But if the prices decline, the increased interest pushes them even lower.

Discussion

Digital currencies are new economic instruments with special attributes. Probably the most important one of them is the fact that they have no underlying asset, they are not issued by any government or central bank and they bring no interest or dividends. Despite these facts, these currencies, and namely the *BitCoin* currency, have attracted the public attention due to the unprecedented price surges with possible profits of hundreds percent in just several weeks or months. In this paper, we analyzed the dynamic relationship between the *BitCoin* price and the interest in the currency measured by search queries on *Google Trends* and frequency of visits on the *Wikipedia* page on *BitCoin*. Apart from a very strong correlation between price

level of the digital currency and both the Internet engines, we also find a strong causal relationships between the prices and searched terms. Importantly, we find that this relationship is bidirectional, i.e. not only do the search queries influence the prices but also the prices influence the search queries. This is well in hand with the expectations about a financial asset with no underlying fundamentals. Speculation and trend chasing evidently dominate the *BitCoin* price dynamics.

Specifically, we find that while the prices are high (above trend), the increasing interest pushes the prices further atop. From the opposite side, if the prices are below their trend, the growing interest pushes the prices even deeper. This forms an environment suitable for a quite frequent emergence of a bubble behavior which indeed has been observed for the *BitCoin* currency. We believe that the paper will serve as a starting point of the research line dealing with statistical properties, dynamics and bubble-burst behavior of the digital currencies as these provide a unique environment for studying a purely speculative financial market.

Methods

Data. Time series have been obtained from http://www.google.com/trends for *Google Trends*, http://stats.grok.se for *Wikipedia* and http://www.bitcoincharts.com for *BitCoin*. Note that the *Google Trends* series are normalized (so that the maximum value of the series is equal to 100) and rounded whereas the *Wikipedia* series provide the actual number of visits for the given day. For the *BitCoin* prices, we focus on the exchange rate with the USD at Mt. Gox platform as this provides the most liquid market. For the fact that *Google Trends* series are available only at the weekly frequency, we had to reconstruct the weekly series (with a same definition of the

week) for the *BitCoin* prices. The weekly *BitCoin* prices are taken as an average of the daily closing prices of the specific weeks. The analyzed period ranges between 1.5.2011 and 30.6.2013 due to illiquidity of the market in the period before (see Fig. 1 and the main text).

For the purposes of distinguishing between the positive and negative feedbacks for *BitCoin* prices, we create a pair of series – Q_t^+ and Q_t^- – defined as $Q_t^+ = Q_t \mathbb{I}_{\left(P_t - \frac{1}{N}\sum_{i=1}^{N} P_{t-i+1}\right) > 0}$ and $Q_t^- = Q_t \mathbb{I}_{\left(P_t - \frac{1}{N}\sum_{i=1}^{N} P_{t-i+1}\right) \leq 0}$ where Q_t is the search frequency at time t and $\mathbb{I}_\bullet$ is an indicator function equal to 1 if the condition in $\bullet$ is met and 0 otherwise, and N is a number of periods taken into consideration for the moving average. For the *Google Trends* series, we use $N = 4$, i.e. 4 weeks (a trending month), and for the *Wikipedia* series, we utilize $N = 7$, i.e. 7 days (a trading week), due to the different frequency sampling. These two variables serve as a proxy for the search-term activity connected with the positive (Q_t^+) and the negative (Q_t^-) feedback.

Stationarity tests. For testing stationarity, we utilize the Augmented Dickey-Fuller test (ADF)[15] and the KPSS test[14]. ADF has a null hypothesis of a unit root ($d = 1$) against the alternative of no unit root ($d < 1$) whereas KPPS has a null of stationarity ($d = 0$) against an alternative of a unit root ($d = 1$). Using the pair of tests, we are able to identify whether the tested series is stationary or not.

If both analyzed series contain a unit root, we can test them for the cointegration. If both series are stationary, we can utilize the vector autoregression (VAR) framework.

Cointegration. We say that two series $\{x_t\}$ and $\{y_t\}$ are cointegrated $CI(d, b)$ if they are both integrated of the same order d and there exists a linear combination of the two series which is integrated of order $d - b$. The standard cointegration is based on $CI(1, 1)$ relationship, i.e. series $\{x_t\}$ and $\{y_t\}$ contain a unit root (they are both $I(1)$) and there exists $u_t = y_t - \alpha - \beta x_t$ which is $I(0)$, i.e. stationary with short memory[26,27].

If the series are cointegrated, the long-term equilibrium relationship is characterized by

$$y_t = \alpha + \beta x_t + u_t. \tag{1}$$

As long as the series are cointegrated, the parameters can be super-consistently estimated using the simple OLS estimator[28]. The lagged residual series $\{\hat{u}_{t-1}\}$ is called the error-correction term and is interpreted as a deviation from the long-term equilibrium.

To test for the cointegration relationship, we use two Johansen tests[25] – the trace test and the maximum likelihood test. If the series are found to be cointegrated $CI(1, 1)$, the error-correction model (ECM) or the vector error-correction model (VECM) is standardly applied. If the analyzed series are not cointegrated, we need to proceed with the vector autoregression applied on the first differences of the originally used series.

Vector autoregression. Vector autoregression is a standard procedure for analyzing (ideally causal) relationship between multiple series[29,30]. In a case of the pair of series $\{x_t\}$ and $\{y_t\}$, the vector autoregression of order p (VAR(p)) is written as

$$\Delta x_t = \alpha_1 + \sum_{i=1}^{p} \beta_{1i} \Delta x_{t_i} + \sum_{i=1}^{p} \gamma_{1i} \Delta y_{t_i} + \varepsilon_{1t} \tag{2}$$

$$\Delta y_t = \alpha_2 + \sum_{i=1}^{p} \beta_{2i} \Delta y_{t_i} + \sum_{i=1}^{p} \gamma_{2i} \Delta x_{t_i} + \varepsilon_{2t} \tag{3}$$

with possibly correlated disturbances $\{\varepsilon_{1t}\}$ and $\{\varepsilon_{2t}\}$ and lag p selected according to some measure, usually an information criterion, such as the Akaike Information Criterion (AIC), Hannan-Quinn Information Criterion (HQIC) and Schwarz Information Criterion (SIC). Assuming that series $\{x_t\}$ and $\{y_t\}$ are $I(1)$, their first differences $\{\Delta x_t\}$ and $\{\Delta y_t\}$ are $I(0)$ and thus stationary so that the system can be easily estimated using either the ordinary least squares or maximum likelihood procedures. Parameters β_1, β_2, γ_1 and γ_2 are themselves not as important as the statistical inference based on them, for our purposes mainly the Impulse-Response analysis. Impulse-Response analysis is based on a vector moving average representation of VAR and it shows what is the reaction of one variable to a unit shock in some other variable and how the effect vanishes in time. For details, see Refs. 29–32.

Vector error-correction model. Vector error-correction model (VECM) is a generalization of the vector autoregression which incorporates the long-term corrections so that both short-term and long-term dynamics can be studied. For cointegrated $CI(1, 1)$ series, we have (VECM(q)) with q lags written as

$$\Delta x_t = \delta_1 + \sum_{i=1}^{q} \theta_{1i} \Delta x_{t_i} + \sum_{i=1}^{q} \kappa_{1i} \Delta y_{t_i} + \lambda_1 \hat{u}_{t-1} + \epsilon_{1t} \tag{4}$$

$$\Delta y_t = \delta_2 + \sum_{i=1}^{q} \theta_{2i} \Delta y_{t_i} + \sum_{i=1}^{q} \kappa_{2i} \Delta x_{t_i} + \lambda_2 \hat{u}_{t-1} + \epsilon_{2t} \tag{5}$$

where parameters θ_i and κ_i control for the short-term dynamics and λ_i represent the error-corrections to the long-term cointegration relationship from Eq. 1. VECM(q)

framework allows for a similar Impulse-Response analysis as the VAR framework. The main difference lays in the fact that the Impulse-Response in the VAR framework illustrates immediate responses whereas in the VECM framework, the permanent shifts in the studied variables are examined[26,27,32].

1. Gordon, M. J. *The Investment, Financing, and Valuation of the Corporation* (Irwin, R. D. & Homewood, I. L. 1962).
2. Krugman, P. & Obsfeld, M. *International Economics* (Pearson Education, Inc., Boston, MA, 2009).
3. Reinert, K. A. & Rajan, R. S. (eds.)*The Princeton Encyclopedia of the World Economy I* (Princeton University Press, Princeton, NJ, 2009).
4. Levi, M. D. *International Finance* (Routledge, Abingdon, 2005).
5. Feenstra, R. C. & Taylor, A. M. *International Macroeconomics* (Worth Publishers, London, 2008).
6. Mondaria, J., Wu, T. & Zhang, Y. The determinants of international investment and attention allocation: Using internet search query data. *J. Int. Econ.* **82**, 85–95 (2010).
7. Preis, T., Reith, D. & Stanley, H. E. Complex dynamics of our economic life on different scales: insights from search engine query data. *Philos. Trans. R. Soc. A-Math. Phys. Eng. Sci.* **368**, 5707–5719 (2010).
8. Drake, M. S., Roulstone, D. T. & Thornock, J. R. Investor information demand: Evidence from google searches around earnings announcements. *J. Account. Res.* **50(4)**, 1001–1040 (2012).
9. Preis, T., Moat, H. S., Stanley, H. E. & Bishop, S. R. Quantifying the advantage of looking forward. *Sci. Rep.* **2**, 350 (2012).
10. Preis, T., Moat, H. S. & Stanley, H. E. Quantifying trading behavior in financial markets using Google Trends. *Sci. Rep.* **3**, 1684 (2013).
11. Moat, H. S. *et al.* Quantifying wikipedia usage patterns before stock market moves. *Sci. Rep.* **3**, 1801 (2013).
12. Kristoufek, L. Can Google Trends search queries contribute to risk diversification? *Sci. Rep.* **3**, 2713 (2013).
13. Nakamoto, S. Bitcoin: A peer-to-peer electronic cash system. http://bitcoin.org/bitcoin.pdf, visited on 11 November 2013.
14. Kwiatkowski, D., Phillips, P., Schmidt, P. & Shin, Y. Testing the null of stationarity against the alternative of a unit root: How sure are we that the economic time series have a unit root? *J. Econom.* **54**, 159–178 (1992).
15. Dickey, D. & Fuller, W. Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.* **74**, 427–431 (1979).
16. Bahmani-Oskooee, M. Export growth and economic growth: an appliation of cointegration and error-correction modeling. *J. Dev. Areas* **27**, 535–542 (1993).
17. Islam, M. Export expansion and economic growth: testing for cointegration and causality. *Appl. Econ.* **30**, 415–425 (1998).
18. Johansen, S. & Juselius, K. Maximum likelihood estimation and inference on cointegration – with applications to the demand for money. *Oxf. Bull. Econ. Stat.* **52**, 169–210 (1990).
19. Miller, S. Monetary dynamics: An application of cointegration and error-correction modeling. *J. Money Credit Bank.* **23**, 139–154 (1991).
20. Hakkio, C. & Rush, M. Market efficiency and cointegration: an application to the sterling and deutschemark exchange markets. *J. Int. Money Finan.* **8**, 75–88 (1989).
21. Pedroni, P. Purchasing power parity tests in cointegrated panels. *Rev. Econ. Stat.* **83**, 727–731 (2001).
22. Narayan, P. The saving and investment nexus for China: Evidence from cointegration tests. *Appl. Econ.* **37**, 1979–1990 (2005).
23. Masih, A. & Masih, R. Energy consumption, real income and temporal causality: results from a multi-country study based on cointegration and error-correction techniques. *Energy Econ.* **18**, 165–183 (1996).
24. Lee, C.-C. Energy consumption and GDP in developing countries: A cointegrated panel analysis. *Energy Econ.* **27**, 415–427 (2005).
25. Johansen, S. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models* (Oxford University Press, Oxford, NY, 1995).
26. Engle, R. F. & McFadden, D. L. (eds.) *Handbook of Econometrics, Vol. IV* (Elsevier, Amsterdam, 1994).
27. Hatanaka, M. *Time-Series-Based Econometrics: Unit Roots and Co-Integration* (Oxford University Press, Oxford, NY, 1996).
28. Engle, R. F. & Granger, C. W. J. Co-integration and error correction: Representation, estimation, and testing. *Econometrica* **55**, 251–276 (1987).
29. Sims, C. A. Macroeconomics and reality. *Econometrica* **48**, 1–48 (1980).
30. Lütkepohl, H. *Applied Time Series Econometrics* (Springer, Berlin, 2005).
31. Hamilton, J. D. *Time Series Analysis* (Princeton University Press, Princeton, NJ, 1994).
32. Enders, W. *Applied Econometric Time Series* (John Wiley & Sons, Hoboken, NJ, 2003).

Author contributions

L.K. solely wrote the main manuscript text, prepared the figures and reviewed the manuscript.

Additional information

Competing financial interests: The authors declare no competing financial interests.