



Say anything you want about me if you spell my name right: the effect of Internet searches on financial market

Aleksandra Rutkowska¹ · Agata Kliber¹

Published online: 2 January 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

In this paper, we focus on the influence of attention from investors on the dynamics of different financial assets. Investor attention is measured by the number of Google searches for three types of assets: stocks, gold and cryptocurrencies. We analyse the daily price of: the Standard & Poor's 500 (S&P 500), the Russell 2000 Index (RUT), Bitcoin and the gold spot price in USD. The study covers a period of 5 years: 2013–2018. According to the results, the prices of different assets react differently to changes in the level of investor attention (and vice versa), and this relationship changes over time. Bitcoin seems to be the most sensitive to the changes in investor attention and changes in its dynamics appear to influence investor attention the most. Changes in the volatility and trade volume of stocks occur simultaneously with changes in investor interest. However, during periods of increased stock volatility, we can observe increased investor interest in gold over any other asset. The price of gold itself has proven itself to be most immune to changes in investor attention.

Keywords Behavioural finance · Google trends · Cryptocurrencies · GARCH models · Causality

1 Introduction

Most economic studies of the last century are influenced by the efficient-market hypothesis (EMH). However, some studies initiated in the late 1960s, together with economists and psychologists such as Daniel Kahneman, Amos Tversky, Richard Thaler, and Paul Slovic, linked the imperfections in financial markets to a combina-

✉ Aleksandra Rutkowska
aleksandra.rutkowska@ue.poznan.pl
Agata Kliber
agata.kliber@ue.poznan.pl

¹ Department of Applied Mathematics, Poznan University of Economics and Business, Niepodleglosci 10, 61-875 Poznan, Poland

tion of cognitive biases and various errors in reasoning and information processing. It has often been questioned whether it is true that all information is available to everyone at the same time, how such information affects individuals, what the impact on the decision process actually is and if it is fully reflected in capital markets. A review of the literature on behavioural finance can be found in e.g. in Fama (1998) and Shiller (2003).

Because of the Internet, people are becoming increasingly influenced by information which is spreading at a rapid pace and reaching a wider and wider audience—even if such information is not verified or reliable. In this study, we focus on information as measured by the number of Google searches for three types of assets: stocks, gold and cryptocurrencies. In the literature, the number of searches is frequently called search volume index (SVI). Since the literature indicates different strengths of influence of investor sentiment towards large and small companies, we analysed two stock indices: one for small-cap stocks (Russell 2000) and the second one for companies with high capitalization (S&P 500).

The aim of the research is to investigate how returns and the volatility of stocks, gold and Bitcoin are related to SVIs. To accomplish this, we estimated a set of GARCH models with explanatory variables (the respective SVI) and verified their significances. Additionally, we investigated the direction of causality between prices and SVIs, as well as between the volume of trade and SVIs, applying the Toda–Yamamoto test of non-causality.

The contribution of the paper is as follows. Although, similarly to previous researchers, we used Google Trends as a measure of the attention, we extended the analysis to include daily data, instead of just weekly ones (see the Literature Review section). Moreover, since we believe that the development of both: Google search popularity, as well as the cryptocurrency market could affect the analysis, we investigated the relationships in two sub-samples: 2013–2015 and 2016–2018. The results differ indeed, with respect to the analysed asset and period. Based on daily-data analysis, we concluded that over the period 2013–2015 stock prices of companies of both high and low capitalization (the S&P 500 and the Russell 2000 index, respectively) influenced investor attention. The opposite relationship was also confirmed: an increase in investor interest led to an increase in prices. However, starting from the beginning of 2016, no dependence between prices and the number of searches was found. In the case of Bitcoin, feedback relationships exist in the whole period. Eventually, the number of Google searches for the keywords *gold price* were influenced by the price of gold, regardless of the sample tested, but no inverse relationship was found.

The remainder of this paper is organized as follows. In Sect. 2 we present our research hypotheses. Section 3 is devoted to the literature, where different attention proxies are presented. We also present state-of-the art knowledge about the relationship between investor attention and the dynamics of stocks, commodities and cryptocurrencies. In Sect. 4, we briefly comment on the data and the methodology of our study, while in the sections following that we present and discuss our results.

2 Research hypotheses

Cryptocurrencies are controversial digital assets designed as a medium of exchange. The question of whether Bitcoin is a medium of exchange or an asset or a new digital gold equivalent is still open (see e.g. Dyrberg 2015; Bouri et al. 2017 or Baur et al. 2018). In this study, we wanted to check whether cryptocurrencies react to the growth of interest more than gold or stocks.

We formed the following hypotheses:

- Hypothesis 1: The number of Google searches is a measure of investor attention and is interrelated with changes of the price and volume of assets. No matter what the sentiment of the search is, a growth of interest should be accompanied with an increase in the returns and volatility of the prices of respective assets.
- Hypothesis 2: A growth of interest should also be accompanied with an increase in changes and volatility of the volume of respective assets.
- Hypothesis 3: An increase in the volatility of stock prices occurs simultaneously with an increase in interest in alternative forms of investment (gold, Bitcoin) which may be perceived as safe-haven assets.

The last sentence may require a word of explanation. There is a general consensus in the literature that gold is a proper safe-haven asset. Many researchers also claim that Bitcoin may play a similar role as well. Some authors have in fact confirmed that Bitcoin bears some resemblance to gold (Dyrberg 2015), and it has safe-haven properties in some sub-periods (Shahzad et al. 2019). Yet, others have come to very different conclusions. A lack of similarities to gold and the lack of any ability of Bitcoin to serve as a safe-haven asset was proved by Klein et al. (2018) or Baur et al. (2018). Moreover, Symitsi and Chalvatzis (2018) note, that the decrease in the overall portfolio risk due to the low correlation of Bitcoin with other assets is not offset by its high volatility. Furthermore, Kliber et al. (2019) claim, that Bitcoin can act as a safe haven only in a crisis-driven country and when investment in local currency is taken into account. Thus, as there is no clear consensus in the literature as to whether Bitcoin can or cannot be perceived as a safe-haven asset, we decided to verify if Bitcoin does indeed have the properties of a safe-haven asset.

To verify these hypotheses, we estimated a set of GARCH-type models for changes of prices and volume with explanatory variables. The variables are the changes of search intensity for respective keywords. We were interested whether the explanatory variables are significant and the sign of the respective coefficient.

Eventually, we were interested in the **direction** of the interrelations: does an increase in investor attention cause an increase in asset prices or does an increase in assets prices cause an increase in investor attention? Alternatively, is it volume that affects attention? Thus, our fourth hypothesis is related to attention theory:

- Hypothesis 4: An increase in investor attention causes an increase in asset prices and volume.

To verify the last hypothesis we applied the Toda–Yamamoto causality test (Toda and Yamamoto 1995) and used Wald statistics to determine the direction of the causality. Note that the first two hypotheses may seem to overlap with the last one. However,

in the first two we measured co-occurrence of the phenomena, while in the last one, we were interested in causality. Moreover, in the first three hypotheses, we investigated stationary returns and changes of volume, while in the fourth one we concentrated on non-stationary prices and levels of volume.

3 Literature review

Kahneman (2018) summarized studies on a variety of aspects of attention, including divided attention, task interference, and the role of perception. According to his filter theory, attention sets a filter to select a certain class of stimuli and to reject all others. Individuals have a limited amount of attention that they can devote to investing.

Attention can affect the trading behaviour of investors in two distinct ways. On the one hand, directing too little attention to important information can result in a delayed reaction to important information. On the other hand, devoting too much attention to information (perhaps old or irrelevant) can lead to an overreaction (Barber and Odean 2013). Moreover, an excess of investment opportunities for analysis creates a need to filter them.

The phenomenon of attention-driven behaviour can be observed in historical data. Barber and Odean (2007), based on trading and position records for the investments of 78,000 households from January 1991 to December 1996 and news from the Dow Jones News Service, confirmed that individual investors exhibit attention-driven buying. Facing many possibilities, investors solve the problem of choice by considering only those stocks that have recently caught their attention. On high-volume days and when stocks are in the news, they are net buyers, following both extremely negative and extremely positive one-day returns. Attention-driven buying is similar when large capitalization stocks are considered, as well as in the case of the small ones.

Such phenomenon can also be observed in experimental studies. Libby et al. (2002) dealt with the problem of the manipulation of information in the context of the attention-driven buying. They showed that when an investor faces many alternatives, options that attract an investor's attention are more likely to be considered and to be chosen. Options that do not attract an investor's attention are often ignored. Thus, if online information can switch our attention filter and focus, it exposes us to potential manipulation.

In the context of access to information, researchers differentiate between informed and uninformed investors and investigate the phenomenon of so-called insider trading (e.g. Jaffe 1974; Seyhun 1986). Some researchers, using historical data, have investigated the possible implications of insider trading for the insiders themselves, outsiders and other stock market observers, such as analysts. According to the stealth trading hypothesis (Barclay and Warner 1993), in which informed traders may try to hide their information, it appears that the less informed traders can profit from the insiders' disclosures by imitating their decisions (see e.g. Gurgul and Majdosz 2007 or Bedowska-Sojka 2014). Other researchers have studied the phenomenon through experimental studies (see e.g. Alfarano et al. 2006; Huber et al. 2008; Marsden and Tung 1997 or Hornung et al. 2015). Their results provide insights into whether and how insiders try to stay undetected and how their profits are accumulated. Such experiments

also verify how the value of information changes over time. In all cases, the authors confirmed that access to proper information is a key factor in taking successful investment decisions. Therefore, investors look for information from a variety of possible sources. Naturally, we can assume that nowadays the best source is the Internet.

3.1 Attention and sentiment proxies

Nowadays, most activities related to investing and collecting data take place online. Not only is the Internet the dominant source of information for traders, but also the information that the traders themselves leave on the Internet becomes a source for other market participants. For instance, analysing the search intensity for particular stock in search engines, or discussions through social media, market analysts can infer so called investor sentiment.

Tetlock (2007) measured the interactions between the media and the stock market using daily content from a popular Wall Street Journal column. They showed that high media pessimism predicts a downward pressure on market prices followed by a reversion to fundamentals, while an unusually high or low pessimism predicts a high market trading volume. Bollen et al. (2011) investigated whether public mood, measured by posted tweets, can predict values of the Dow Jones Industrial Average (DJIA). Among the seven observed mood patterns, only some were Granger-caused by the DJIA and matched shifts in DJIA values that occurred 3 to 4 days later (mainly mood labelled as “calm” from the Google-Profile of Mood States). Eventually, Moat et al. (2013) noticed a significant increase in the number of views of Wikipedia articles relating to companies or other financial topics before a decline of stock market prices. Eventually, Green et al. (2019) examined Virtual Globe technology and showed how the arrival of unprecedented types of web-based information enhances the ability to forecast and can lead to significant, measurable economic benefits.

Since 2006, a large and interesting source of data is Google Trends. Google Trends is a public web facility that shows how often a search-term is entered into the Google search engine. Obviously, the growth of popularity of this search engine over the last 20 years is extraordinary. In 1998, the average number of queries per day was 10,000, but in 2006 it was 10,000 search queries per second. Statistics show that in 2018 it was 63,000 search queries per second (around 5.5 billion per day). As a relatively new and big data source, this tool has also been recently widely used as a sentiment proxy (see e.g. Rutkowska and Kliber 2018) and to forecast and analyse financial markets, but with mixed results.

3.2 Google search intensity and stock markets dynamics

The first studies on the impact of searches on stocks concentrated on developed markets. Joseph et al. (2011) examined online stock ticker searches in a sample of S&P 500 firms over the period 2005–2008 and confirmed that over a weekly period online search intensity reliably predicted abnormal stock returns and trading volume. Analogous conclusions came from Bank et al. (2011), who analysed the German stock market. The authors showed that an increase in search queries was linked to a growth

of stock liquidity and temporarily higher returns. This finding was partially confirmed a few years later by Bijl et al. (2016), who found that large numbers of searches were significant but weak predictors of future excess returns on German stock markets. Da et al. (2011)—on the contrary—concentrated on the small stock index: the Russell 3000. The authors proposed a direct measure of investor attention using search data in Google and found that an increase in this measure predicted an increase in stock prices in the following 2 weeks and a possible price reversal within one year. Preis et al. (2013) took into account the dynamics of the DJIA. They built a strategy where a market portfolio was bought or sold, based on Google search volumes. With the keyword *debt* this strategy outperformed the market index by 326%. Dimpfl and Jank (2016) concentrated not on prices but on volatility of the DJIA and showed that a rise in the number of searches today is followed by an increase in volatility tomorrow. The authors improved volatility forecasts in-sample.

In all the studies mentioned above, developed markets were analysed. Much less attention has been placed on emerging ones. The results presented so far indicate, however, that also in this case causal relationships can be found. For instance, Tantaopas et al. (2016) analysed a group of developed and developing countries to verify the possible influence of search intensity on their stock markets. The researchers checked for Granger causality between Google searches and selected Asia-Pacific equity markets based on weekly data between January 2004 and December 2014. According to their findings: the main index return of Hong Kong, Japan, Korea, China, and India Granger-caused the search volume. On the other hand, Australia and New Zealand were the countries where search intensity moved index returns. When volatility and search volume index were considered, in most countries (Japan, India, and Malaysia) two-way relationships existed: return volatility was influenced by search intensity, and vice versa. Moreover, the variation of returns Granger-caused search volume in three developed markets (Australia, Korea, and New Zealand). For China, the results were different, as the changes in search index influenced changes in the volatility of its representative index. The trading volume and SVI exhibited the least numbers of statistically significant causality among the other pairs. In three countries (Korea, China, and India), a change in their trading volumes Granger-caused changes in corresponding Google searches, but not vice versa.

There is also the work of Nguyen et al. (2019), who analysed 5 emerging markets (Indonesia, Malaysia, Philippines, Thailand, and Vietnam) over the period 2009–2016 and found that increases in Google search volume have significant negative impacts on stock returns in the case of Philippines, Thailand, and Vietnam. This suggests that investors might be more sensitive to bad news than good news in their investment decisions.

3.3 Google search intensity and the dynamics of commodities

All the examples cited above proved that the relationships between SVIs and the dynamics of stock markets exist (but can be different in their nature). What is interesting, is the fact that similar dependencies can be found for instance in the case of commodity markets. Let us start from the work of Vozlyublenniaia (2014) who studied

stocks, commodities and treasury bonds. The data were taken weekly and covered the period: January 2004 to December 2012. The author analysed the following indices: the Dow Jones Industrial Average, the S&P 500 index, NASDAQ, the 10-year Treasury Bond index, the Chicago Board Options Exchange (gold), and West Texas Intermediate (crude oil). She showed that changes in investor interest in oil caused changes of oil returns and vice versa. Oil returns influenced search intensity of the DJIA, while DJIA returns affected the search likelihood of oil. Eventually, changes in gold returns influenced the future search intensity for the gold index. In the case of volatility: it caused changes in attention, but attention was less likely to influence volatility. These relationships may have occurred within the same index as well as across indices. Yet another example can be Li et al. (2015), who studied the relationships between SVIs, different trader positions in the Commitments of Traders (COT) reports by the CFTC, and crude oil price changes. The authors found that SVIs significantly drives speculators' positions.

3.4 Influence of internet search activity on Bitcoin price dynamics

The literature on the influence of SVIs on Bitcoin prices is more scarce and has only started to appear recently. In such cases, the analyses were performed also using daily data. One of the first studies was Kristoufek (2015) who showed that there are significant interrelations between the price of Bitcoin and investor interest as measured by Google searches, and that their direction and intensity changes over time. Similar conclusions were reached by Suessmuth (2019) for a later period. Eventually, Abraham et al. (2018) showed that Bitcoin and Ethereum price predictions can be improved with the help of data on tweet volumes and Google trends. The authors proved that it is the volume of searches, not their sentiment, that allows for more accurate predictions of price direction. On the contrary, Aalborg et al. (2019) showed that only predictions of Bitcoin volume—not the returns—can be improved with Google search data.

Our study is similar to the ones cited above in the sense that we utilize Google search data as a measure of attention. Contrary to previous studies, however, we used daily data also for stocks and gold, as the effects of attention from noise traders are more likely to manifest at shorter horizons than one week (e.g. during one day). In this sense, our study is an extension and continuation of the previous ones. We expanded the sample to cover not only stocks but also cryptocurrencies and gold, so that we can extend the inference for alternative investment opportunities. Moreover, we divided the whole 5-year sample into two subsamples to account for the development of the Google search engine as well as the development of cryptocurrencies.

4 Data and methodology

Our data cover the period from 30-01-2013 to 28-01-2018. We analysed the daily prices of:

- the S&P 500, an American stock market index based on the market capitalization of 500 large companies which have common stock listed on the NYSE or NASDAQ,

- the Russell 2000 Index, a small-cap stock market index of the bottom 2000 stocks in the Russell 3000 Index, maintained by the FTSE Russell,
- Bitcoin, a cryptocurrency of the biggest market capitalization,
- Gold Spot US Dollar, the gold price quoted in US Dollars, which is the common yardstick for measuring the value of Gold across the world,

and daily volume of:

- S&P500 index,
- the Russell 2000 index,
- Bitcoin.

As for the volume of Bitcoin, we chose data showing the USD trade volume from the top exchanges according to <https://blockchain.info>. The data were downloaded from the Quandl service.¹

Since the data are non-stationary, the models and calculations were performed on logarithmic returns. Descriptive statistics of logarithmic rates of return are presented in Table 1. What we noticed is that there is an extremely high value of Bitcoin's kurtosis as compared to the other assets. Also, the range of the values taken by Bitcoin returns, as well as the sample variance, are significantly higher. However, when we compared the volume of trade of Bitcoin and stocks, we can see that it was again more volatile in the case of Bitcoin (measured by standard deviation), but the kurtosis is the lowest.

By analysing Figs. 1, 2, 3 and 4, in which we present the prices of the analysed data, together with the search intensity, we can easily observe that there are several moments of possible co-movement. Such an example is August 2015, when the S&P 500 closed about 1.3 per cent lower, after rallying nearly 3 per cent earlier, which was the biggest reversal to its downside since October 29, 2008. The movements of Bitcoin prices and volume probably corresponded to the major events in the market, such as the issuing of the first regulatory guidelines for *decentralised virtual currencies* by the Financial Crimes Enforcement Network in 2013 or the shutting down of the Silk Road website by the FBI which seized roughly 26,000 BTC from various accounts (Autumn, 2013).

4.1 Keywords choice

As a measure of attention, we used the daily number of Google keyword searches, called *Google Trends*. We used three keywords closely related to a number of chosen assets: *stock*, *bitcoin*, *gold price*.

Choosing a keyword is always problematic and subjective. We tested between 5–20 synonyms and words related to each type of asset. Eventually, we chose the most popular ones. The examples of the queries can be seen in Figs. 5, 6 and 7, while a list of synonyms is in Table 2. The queries were limited to the United States and its finance categories.

Google Trends adjusts search data to make it easier to compare keywords. Each data point is divided by the total number of searches in the associated geographic region

¹ <https://www.quandl.com/data/BCHAIN/TRVOU-Bitcoin-USD-Exchange-Trade-Volume>.

Table 1 Descriptive statistics of log-returns of the analysed assets and changes of their volume

	BTC	Gold	S&P500	Russel 2000
Price				
Mean	0.0035	-0.0002	0.0005	0.0005
Standard error	0.0011	0.0003	0.0002	0.0003
Median	0.0025	-0.0001	0.0006	0.0011
Standard deviation	0.0486	0.0097	0.0075	0.0099
Sample variance	0.0024	0.0001	0.0001	0.0001
Kurtosis	15.8382	8.6956	2.8784	0.9398
Skewness	-0.7871	-0.7402	-0.4411	-0.3653
Range	0.8371	0.1411	0.0785	0.0713
Minimum	-0.4783	-0.0924	-0.0402	-0.0398
Maximum	0.3588	0.0487	0.0383	0.0315
Volume				
Mean	0.0036		0.0000	-0.0001
Standard error	0.0135		0.0075	0.0051
Median	-0.0358		-0.0032	-0.0034
Standard deviation	0.5749		0.2655	0.1812
Sample variance	0.3306		0.0705	0.0328
Kurtosis	0.8225		6.4003	16.8135
Skewness	0.3545		-0.0633	-0.1611
Range	4.4750		2.4202	3.4686
Minimum	-1.7749		-1.1952	-1.7777
Maximum	2.7001		1.2250	1.6909

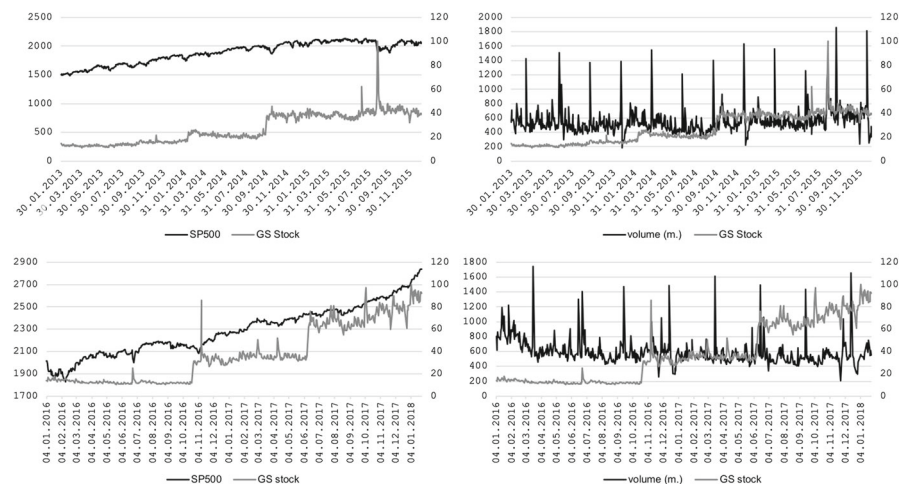


Fig. 1 SP500 price (left panel, left axis, black solid line) and volume in millions (right panel, left axis, black solid line) versus number of searches for *stock* from Google Trends (right axis, grey solid line) over the period from 2013-01-31 to 2015-12-31 (upper panel) and from 2016-01-01 to 2018-01-28 (lower panel)

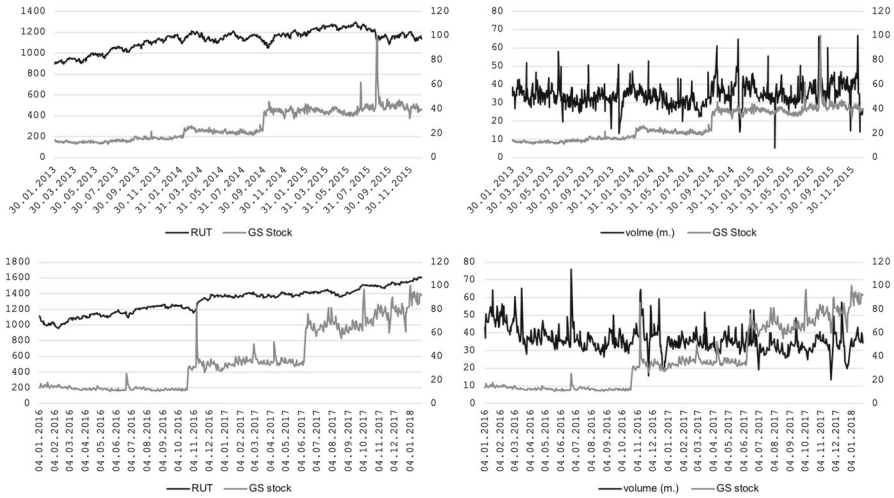


Fig. 2 Russell 2000 price (left panel, left axis, black solid line) and volume in millions (right panel, left axis, black solid line) versus number of searches for *stock* from Google Trends (right axis, grey solid line) over the period from 2013-01-31 to 2015-12-31 (upper panel) and from 2016-01-01 to 2018-01-28 (lower panel)

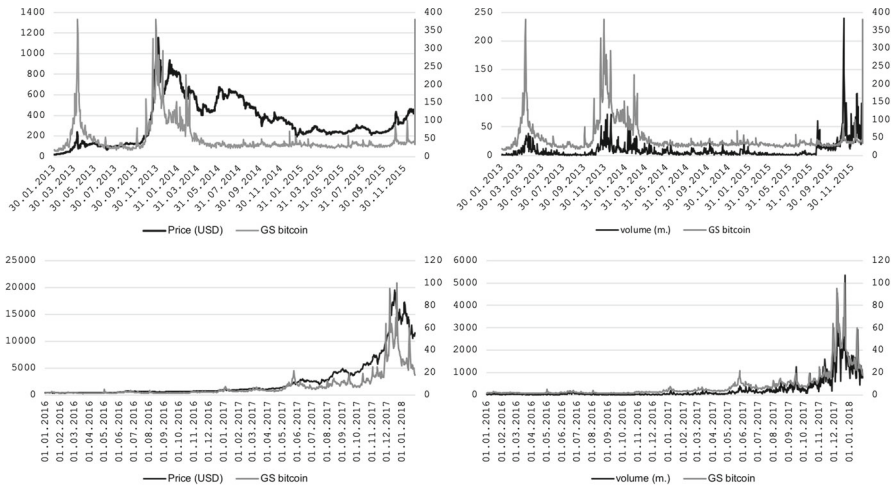


Fig. 3 Bitcoin price (left panel, left axis, black solid line) and volume in millions (right panel, left axis, black solid line) versus number of searches for *bitcoin* from Google Trends (right axis, grey solid line) over the period from 2013-01-31 to 2015-12-31 (upper panel) and from 2016-01-01 to 2018-01-28 (lower panel)

and time range to determine its relative popularity. The obtained values are scaled to the range of 0-100. If the number of searches is low, it is displayed as zero. Google Trends omits identical queries entered by the same user at short intervals.

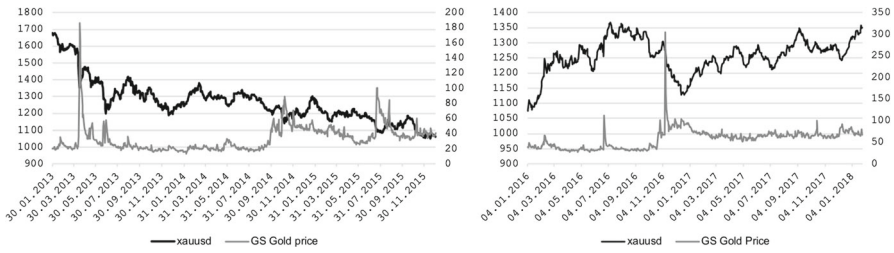


Fig. 4 Gold price (left axis, black solid line) versus number of searches for *gold price* from Google Trends (right axis, grey solid line) over the period from 2013-01-31 to 2015-12-31 (upper panel) and from 2016-01-01 to 2018-01-28 (lower panel)

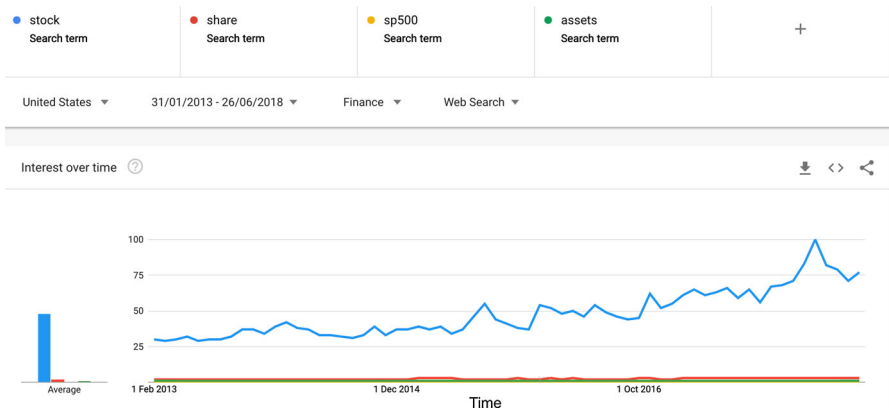


Fig. 5 Comparison of the search volume index for keywords related to the stock market. *Source:* Google Trends

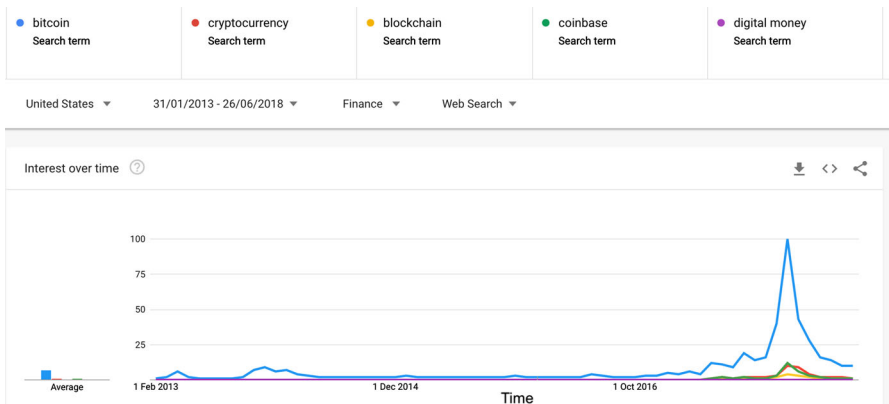


Fig. 6 Comparison of the search volume index for keywords related to the bitcoin market. *Source:* Google Trends

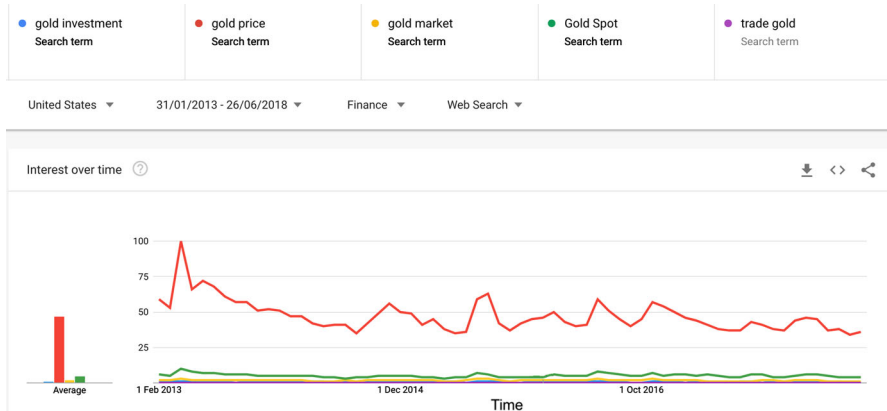


Fig. 7 Comparison of the search volume index for keywords concerning gold market. *Source:* Google Trends

Table 2 List of tested words

Gold	S&P 500 Index / Russell 2000 Index	Bitcoin
gold bars,	Cboe BZX Exchange,	Binance,
gold coin,	equity,	bitbay,
gold futures,	NASDAQ,	bitcoin,
gold price,	assets,	Bitfinex,
spot gold,	market assets,	Bithumb,
XAU	market index,	Bitstamp,
	NYSE,	Bittrex,
	Russell 2000 Index,	blockchain,
	S&P 500 Index,	BTC,
	shareholding,	coinbase,
	shares,	Coinbase,
	sp500,	cryptocoin,
	stock exchange,	cryptocurrency,
	stock investment,	digital currency exchange,
	stock market,	digital currency,
	stock market,	Huobi,Kraken
	stocks	OKEx,Upbit,
		XBT

4.2 Sample split

As already mentioned, our study covers data from 30-01-2013 to 28-01-2018. We started our sample in 2013 when the number of Internet users worldwide amounted to 2631 million and Google’s search engine reached 114.7 billion searches, i.e. its market share was equal to 65.2% (the share of other search engines did not exceed

10%). The sample was divided into two sub-samples: from 2013-01-31 to 2015-12-31 and from 2016-01-01 to 2018-01-28. The division of the sample is justified by the fact that the digital world in 2013 was not comparable to the one in 2018. Internet users went mobile (in 2017, 237.2 million people accessed the Internet through their mobile phones²), the number of Google searches in 2018 exceeded 3.5 billion searches per day,³ and Bitcoin's market capitalization had increased from approximately 0.04 billion U.S. dollars in the first quarter of 2012 to approximately 237.62 billion U.S. dollars in the fourth quarter of 2017.⁴

The particular date to split the time series was chosen based on an analysis of the structural breaks of the Bitcoin time series. In the first step of our research, we estimated all the models of the whole sample. In all the cases, the models were well-fitted, apart from the case of Bitcoin. In the latter case, the model did not pass the Nyblom test of the stability of coefficients. Such a result suggests a possible structural change in the data. Therefore, we decided to look for possible breakpoints in the dynamics of Bitcoin's price. In the test of Bai and Perron (2003) one can choose the preferred number of breakpoints. In our case, the algorithm suggested a maximum of 6 breakpoints. However, this would result in an overly short time series for analysis (i.e. the estimates of the GARCH coefficients would be unstable and thus it would not be reasonable to make any conclusions based on such models). Therefore, we decided to divide the series into two optimal segments. The algorithm indicated 2016 as the moment of division. The length of the data sample in the first sub-period is 1066 days, and in the second one: 759 days. The maximum time frame for daily SVI data is 270 days. For a longer time span only weekly data are available, therefore the data have been scaled to combine it with daily data. To collect and re-scale data we use API for Google Trends in Python available on: <https://github.com/GeneralMills/pytrends>.

4.3 Possible monday effect in keywords search

The raw market data of our analysis are shown in Figs. 1, 2, 3 and 4. Market data for stocks and gold are not quoted during weekends, therefore we had to deal with this problem in some way. In financial markets, some assets may exhibit the so-called: Monday effect, i.e. on Mondays, stock returns are often significantly lower compared to those of the preceding Friday. Various studies (Chiah and Zhong 2019; Caporale and Plastun 2019; Keef et al. 2009) show that such an effect exists in the case of stocks but is not observed in the case of gold or Bitcoin. In our study, we checked two options: omitting weekend search data versus their averaging. The differences between the two approaches were not significant, which may suggest that the Monday effect does not exist (neither was present in the analysed period). We would like to point out that this approach is in line with the work of others, e.g. Aalborg et al. (2019). A detailed investigation of this issue is, however, out of the scope of this study. In this paper, we make a simplifying assumption about the lack of the Monday effect and show the results without including searches made during weekends.

² <https://www.statista.com/statistics/275591/number-of-mobile-internet-user-in-usa/>.

³ <http://www.internetlivestats.com/google-search-statistics/>.

⁴ <https://www.statista.com/statistics/377382/bitcoin-market-capitalization/>.

4.4 Methodology

As the financial time series are characterised not only by linear, but above all by non-linear relationships, we decided to estimate a series of GARCH models (see the “Appendix” for details) to describe their full dynamics. Due to the non-stationarity of the data, the models are estimated on log-returns of asset prices. To account for the possible influence of the search intensity on the assets’ dynamics, we included explanatory variables into conditional mean and variance equations.

Thus, to test the first two hypotheses (Hypothesis 1 and 2): *No matter what the sentiment of the searches is, a growth of the interest and an increase in returns and the volatility of price and volume of respective assets should be observed simultaneously*, we fit an ARMA-GARCH-type model to the daily logarithmic returns of stock, Bitcoin and gold prices with search intensities for respective assets. We did the same for the trade volume of stocks and Bitcoin. If the coefficients were significant, we concluded that investor attention is significantly linked to the assets’ prices or trade dynamics. We expected that the sign of the coefficient of explanatory variables would be positive (the growth of interest is related to a growth of returns/volume changes, and also volatility).

To verify the third hypothesis (Hypothesis 3): *An increase in the volatility of stock prices occurs simultaneously with an increase in interest in alternative forms of investment (gold, Bitcoin) perceived as safe-haven assets*, we extended the set of explanatory variables in the model of stock (S&P and RUT) volatility, to the number of searches for additional assets. We expected that an increase in stock volatility would occur simultaneously with an increase in investor attention for alternative forms of investments. The justification is as follows: an increase in volatility in mature financial markets is associated more frequently with an inflow of bad news—not good news (see: Black 1976). In the case of a decline in stock prices (and a growth of volatility), investors should look for alternatives—e.g. safe-haven assets. Therefore, we should observe positive relationships with stock volatility growth and increased interest for other investment opportunities (as measured by search intensity).

In all the cases, we chose the best model based on its ability to explain linear and non-linear relationships in the data, as well as on the significance of its coefficients. We also compared information criteria. In some models, we decided to leave some insignificant explanatory variables, if such model had better characteristics.

Eventually, to verify the fourth hypothesis (Hypothesis 4): *An increase in investor attention causes an increase in asset prices and volume*, we performed the non-causality test of Toda and Yamamoto (1995). The application of this test, contrary to the classical Granger one (1969), is not limited to stationary data. As our data are not stationary, we decided to use this approach and also analyse levels of prices and volume. It allowed us to test, whether a change of attention causes a change of asset prices and volume, and vice versa. The conclusions of the tests are based on Wald statistics.

5 Results

In the following subsections, we present the results of the GARCH model estimation (hypotheses 1–3) and the results of the Toda–Yamamoto test for non-causality. In the case of GARCH estimates, we were interested whether the explanatory variables (SVIs) are significant in volatility equations. In the case of the causality test, we checked the significance of Wald statistics.

5.1 Investor attention and the dynamics of asset returns and volume (hypotheses 1 and 2)

In this section we present the results of GARCH model estimations that allowed us to verify the first two hypotheses.

In Table 3 we present the results of the estimation of the GARCH model for the S&P 500. In the upper panel, we present the model for returns, while in the lower—for volume. In the case of returns, the best model was AR(1)-IGARCH(1,1): in the first period with skewed Student distribution, while in the second—with a symmetric one. In the case of volume, in the first period we fit the ARFIMA(1,d,0)-ARCH(1) model, while in the second—AR(2)-ARCH(1), in both sub-periods with skewed Student distribution. When we analysed the significance of the explanatory variables (SVI), we could see that an increase in the number of searches for the keyword *stock* is interrelated with a growth of S&P 500 volatility in both sub-samples (the coefficient for the keyword *stock* denoted in the table by GSStock(V) is positive in both periods). In contrast to Da et al. (2011), Bank et al. (2011) and Joseph et al. (2011) who improved forecasts of S&P returns by taking account of search volume indices, we cannot confirm any statistically significant relationship between returns and SVIs (GSStock(M) was insignificant in the first period, while in the second one its p value amounted to 0.07). However, as already noted, dependencies do exist in volatility, which is in line with the results obtained for weekly data by Dimpfl and Jank (2016) for DIJA.

Additionally, strong and positive relationships in mean can be found between S&P 500 volume and the SVIs (GSStock was positive and significant explanatory variable of conditional mean), which corroborates the results obtained for weekly data by Joseph et al. (2011).

In the case of small caps index returns (see Table 4), in the first period the AR(0)-GARCH(1,1) model with G.E.D. distribution was fit, while in the second: AR(1)-ARCH(4) with the same distribution. The best model for volume in the first period was ARFIMA(0,d,0)-ARCH(1) with symmetric Student distribution, while in the second: ARFIMA(3,d,0)-ARCH(1) with G.E.D. distribution. Again, we were interested in the significance of explanatory variables (SVI). Contrary to the case of the S&P 500, we observed interrelations between investor attention and stock volatility in the second sub-sample only. The volume of the Russell 2000 is also positively interrelated to SVIs. In the first sub-sample, an increase in investor attention goes along with an increase in conditional volatility and the mean of volume process, while in the second one—the mean only.

Table 3 Estimates of the GARCH model for S&P500 returns and volume with *stock* SVI as an explanatory variable

	2013-01-31–2015-12-31				2016-01-04–2018-01-26			
	Coeff.	SD error	t-value	t-prob	Coeff.	SD error	t-value	t-prob
SP500 returns								
μ	0.06	0.02	2.69	0.01	μ	0.02	4.17	0.00
AR(1)	-0.11	0.04	-2.57	0.01	GSStock(M)	0.00	1.83	0.07
ω	0.06	0.02	2.75	0.01	AR(1)	-0.13	-2.69	0.01
GSStock(V)	0.01	0.00	3.55	0.00	ω	0.02	1.74	0.08
ARCH (α)	0.18	0.04	4.56	0.00	GSStock(V)	0.01	3.87	0.00
GARCH (β)	0.72	0.06	12.64	0.00	ARCH (α)	0.16	2.59	0.01
Asymmetry	-0.24	0.06	-3.75	0.00	GARCH (β)	0.79	11.38	0.00
Tail	15.97	8.33	1.92	0.06	Student(DF)	4.08	4.75	0.00
SP500 volume								
μ	0.18	0.13	1.38	0.17	μ	0.60	1.15	0.25
min (M)	-100.19	3.26	-30.73	0.00	min (M)	21.14	6.27	<0.01
plus (M)	101.60	2.15	47.34	0.00	GSStock(M)	0.33	3.54	<0.01
GSStock(M)	0.68	0.09	7.73	0.00	AR(1)	-0.46	-13.20	<0.01
d-Arfima	-0.18	0.05	-3.87	0.00	AR(2)	-0.16	-4.86	<0.01
AR(1)	-0.40	0.05	-7.63	0.00	ω	432.44	183.21	0.02
AR(2)	-0.24	0.05	-4.80	0.00	GSStock(V)	-0.11	2.91	0.97
AR(3)	-0.07	0.04	-1.96	0.05	ARCH(α)	0.99	2.18	0.03

Table 3 continued

	2013-01-31–2015-12-31				2016-01-04–2018-01-26			
	Coeff.	SD error	t-value	t-prob	Coeff.	SD error	t-value	t-prob
ω	302.89	62.20	4.87	0.00				
ARCH(α)	0.22	0.09	2.47	0.01	Asymmetry	0.38	0.08	4.75 <0.01
Asymmetry	0.33	0.06	5.44	0.00	Tail	2.58	0.33	7.80 <0.01
Tail	3.13	0.43	7.27	0.00				

In the table, we present the estimates of the AR-GARCH model for S&P returns (upper panel) and volume (lower panel). AR(1), AR(2),AR(3) denote the respective AR coefficients a_i of a conditional mean specification (see the “Appendix” for details). GSSStock(M) denotes the number of Google searches of keyword *stock*—included in a conditional mean equation, while GSSStock(V)—in a conditional variance equation. The min(M) and max(M) variables are binary values included in the conditional mean equation of volume, taking value 1 in the case of the extreme (exceeding 3 standard deviations) value of volume and 0 otherwise. d -Arfima denotes the d coefficient in ARFIMA specification of conditional mean (see the “Appendix”). ARCH and GARCH denote respectively α and β coefficients from GARCH specification of conditional variance. In the case of S&P returns in the first period, the AR(1)-IGARCH(1,1) model was fit with skewed Student error distribution (Asymmetry and Tail denote respective coefficients of the distribution), while in the second period - AR(1)-IGARCH(1,1) with symmetric Student distribution. In the case of the volume—in the first sub-sample dARFIMA(1,d,0)-ARCH(1) was fit with asymmetric Student distribution, and in the second one: AR(2)-ARCH(1) with asymmetric Student distribution

Table 4 continued

	2013-01-31–2015-12-31					2016-01-04–2018-01-26				
	Coeff.	SD error	t-value	t-prob		Coeff.	SD error	t-value	t-prob	
ARCH(α_1)	0.22	0.08	2.83	0.01	AR(3)	0.10	0.00	326.70	0.00	
Student(DF)	3.30	0.41831	7.879	0	ω	123.94	17.40	7.12	0.00	
					ARCH(α_1)	0.26	0.07	3.57	0.00	
					G.E.D.(DF)	0.89	0.07	11.92	0.00	

In the table we present the estimates of the AR-GARCH model for Russell 2000 returns (upper panel) and volume (lower panel). AR(1), AR(2), AR(3) denote the respective AR coefficients a_i of a conditional mean specification (see the "Appendix" for details). GSStock(M) denotes the number of Google searches for the keyword *stock*—included in a conditional mean equation, while GSStock(V)—in a conditional variance equation. d -Arfima denotes the d coefficient in the ARFIMA specification of the conditional mean (see the "Appendix"). ARCH and GARCH denote respectively α and β coefficients from the GARCH specification of conditional variance. In the case of RUT returns in the first period, the AR(0)-GARCH(1,1) model was fit with generalized error distribution (G.E.D.), while in the second period—AR(1)-ARCH(4) with G.E.D. distribution. In the case of the volume—in the first sub-sample dARFIMA(0,d,0)-ARCH(1) was fit with symmetric Student distribution, while in the second one: dARFIMA(3,d,0)-ARCH(1) with G.E.D. distribution

The results confirm our hypothesis that a growth of investor attention is positively related to the dynamics of prices (here: volatility) and the volume of stocks. Thus, periods of increased volatility and increased investor interest occur at the same time. Moreover, there is a significant relationship between the growth of investor interest and the growth of trade intensity.

Similar relationships with price volatility were confirmed for gold (see 5). In this case, we needed more sophisticated volatility models to explain all the dynamics of gold prices. In the first period, the proper model appeared to be AR(1)-HYGARCH(1, d ,1) with G.E.D. distribution, and in the second one: AR(0)-GARCH(1,1). Based on an analysis of the significance of the SVI, we conclude that an increase in the number of searches for the keyword *gold price* and an increase in gold price volatility happens simultaneously.

Different relationship patterns have been found in the case of Bitcoin (see Table 6). In the first sub-period, the AR(5)-GARCH(1,1) model with G.E.D. distribution was fit: both for returns and volume. In the second sub-period, the AR(0)-GARCH(1,1) with G.E.D. distribution was fit for returns, while AR(5)-GARCH(1,1) with G.E.D. distribution for volume. The coefficients of the conditional mean and variance for the SVI are significant (although small) and positive. Thus, we can conclude that an increased interest in Bitcoin is related to an increase in its price and volatility.

Very interesting results were found for volume. In the first sub-sample, the coefficients for the SVI are positive in the conditional mean and negative in the conditional variance equation. Only in this case is an increase in attention related to an increase in the volume and a decrease in the volatility of volume. This can mean that together with an increase of investor attention, the trade volume increased but the increase was not a rapid one.

In the second sub-sample, the situation is the opposite. Thus, an increase in attention was related to a decrease in the volume and an increase in the volatility of volume. This may suggest that an increase of attention in the second period, overlapped with a growth of nervousness in investment behaviour (an increase in the volatility of volume), but not necessarily with an increase in investment itself.

To summarize—our results suggest that there exist significant relationships between investor attention and price and volume dynamics, but they are of a different nature in the case of different assets. The most prone to changes of investor attention is the price of Bitcoin and its volume (relationships have been found both in both mean and volatility equations, as well as in volume). Changes of investor attention are also related to changes of stock volatility and volume, as well as to the volatility of gold prices.

We would like to stress here the fact that this kind of analysis does not allow us to decide what phenomenon was the cause and what was the consequence—i.e. does an increase in attention contribute to an increase in the price, or is it the other way round: an increase in prices results in a growth of attention? This question can be answered if we included a lagged dependent variable in our models. However, in this part of the study, we wanted to concentrate on the relationships within one day, while the causality analysis is performed in Sect. 4.3 (the fourth hypothesis).

Table 5 Estimates of gold price GARCH model with *Gold Price SVI* as an explanatory variable

	2013-01-31–2018-01-25				2016-01-04–2018-01-26			
	Coeff.	SD error	t-value	t-prob	Coeff.	SD error	t-value	t-prob
Gold Spot US Dollar								
μ	-0.06	0.02	-3.18	0.00	μ	0.03	1.43	0.15
AR(1)	-0.09	0.03	-2.95	0.00	ω	0.01	1.20	0.23
GSGold(V)	0.02	0.00	7.80	0.00	GSGold(V)	0.00	2.09	0.04
d-Figarch	0.00	0.00	6.36	0.00	ARCH(α)	0.01	2.24	0.03
ϕ	0.41	0.18	2.27	0.02	GARCH(β)	0.02	42.99	0.00
β	0.54	0.18	3.05	0.00	G.E.D.(DF)	0.12	12.38	0.00
G.E.D.(DF)	1.22	0.09	13.43	0.00				
log(α)	3.24	0.16	20.77	0.00				

In the table we present the estimates of AR-GARCH model for gold returns. AR(1) denotes the α_1 coefficient of conditional mean specification (see the “Appendix” for details). GSGold(M) denotes the number of Google searches for keyword *gold price*—included in conditional mean equation, while GSGold(V)—in conditional variance equation. In the first sub-period the AR(1)-HYGARCH(1,d,1) model of Davidson Davidson (2004) with G.E.D. distribution was fit: d -Figarch denotes the d coefficient in hyperbolic GARCH specification of conditional variance, while ϕ and β —respective model parameters (see the “Appendix”). In the second sub-period the AR(0)-GARCH(1,1) was fit with G.E.D. distribution. ARCH and GARCH in the right panel denote, respectively, α and β coefficients from the GARCH specification of conditional variance

Table 6 Estimates of the Bitcoin price and volume GARCH model with *bitcoin* SVI as an explanatory variable

	2013-01-31–2015-12-31					2016-01-04–2018-01-26				
	Coeff.	SD error	t-value	t-prob		Coeff.	SD error	t-value	t-prob	
BTC price										
μ	0.00	0.00	3.32	0.00	μ	0.00	0.00	7.17	0.00	
GSBitcoin(M)	0.01	0.00	8.48	0.00	GSBitcoin(M)	0.01	0.00	29.88	0.00	
AR(1)	0.00	0.00	0.57	0.57	$\omega \cdot 10^4$	0.19	0.06	3.01	0.00	
AR(2)	-0.05	0.00	-93.91	0.00	GSBitcoin(V)	0.00	0.00	4.07E ⁰⁷	0.00	
AR(3)	0.03	0.00	27.56	0.00	ARCH(α_1)	0.18	0.042	4.33	0.00	
AR(4)	-0.03	0.00	-19.04	0.00	GARCH(β_1)	0.82	0.03	27.33	0.00	
AR(5)	0.03	0.00	38.62	0.00	G.E.D.(DF)	0.88	0.06	15.75	0.00	
$\omega \cdot 10^4$	0.61	0.33	1.83	0.07						
GSBitcoin(V)	0.00	0.00	5.1 · 10 ⁸	0.00						
ARCH(α_1)	0.14	0.07	1.96	0.05						
GARCH(β_1)	0.83	0.07	11.15	0.00						
G.E.D.(DF)	0.96	0.06	17.04	0.00						
BTC volume										
μ	-18.04	6.58	-2.74	0.01	μ	17.12	0.62	27.81	0.00	
GSBitcoin(M)	0.01	0.01	2.68	0.01	GSBitcoin(M)	-0.01	0.00	-9.05	0.00	
AR(1)	-0.38	0.03	-11.45	0.00	AR(1)	0.63	0.04	16.42	0.00	
AR(2)	-0.35	0.03	-10.71	0.00	AR(2)	0.04	0.04	0.89	0.37	
AR(3)	-0.29	0.03	-9.44	0.00	AR(3)	0.05	0.04	1.13	0.26	
AR(4)	-0.22	0.03	-6.53	0.00	AR(4)	0.09	0.04	2.27	0.02	
AR(5)	-0.18	0.03	-5.56	0.00	AR(5)	0.17	0.04	4.76	0.00	

Table 6 continued

2013-01-31–2015-12-31		2016-01-04–2018-01-26		
	Coeff.	SD error	t-value	t-prob
ω	585.25	53.64	10.91	0.00
GSBitcoin(V)	-0.29	0.07	-4.14	0.00
ARCH(α_1)	0.07	0.03	2.27	0.02
GARCH(β_1)	0.85	0.05	15.88	0.00
G.E.D.(DF)	1.60	0.14	11.43	0.00
	Coeff.	SD error	t-value	t-prob
	0.01	0.00	1.33	0.18
	0.00	0.00	5.37	0.00
	0.02	0.01	1.73	0.08
	0.94	0.03	28.56	0.00
	1.86	0.16	11.37	0.00

In the table, we present the estimates of the AR-GARCH model for Bitcoin returns (upper panel) and volume (lower panel). AR(i) denotes the a_i coefficient of the conditional mean specification (see the "Appendix" for details). GSBitcoin(M) denotes the number of Google searches for the keyword *Bitcoin*—included in a conditional mean equation, while GSBitcoin(V)—in a conditional variance equation. In the first sub-period, the AR(5)-GARCH(1,1) model was fit with G.E.D. distribution: both for returns and volume. In the second sub-period the AR(0)-GARCH(1,1) with G.E.D. distribution was fit for returns, while AR(5)-GARCH(1,1) with G.E.D. distribution for volume. ARCH and GARCH denote respectively α and β coefficients from the GARCH specification of conditional variance

5.2 Alternative investments hypothesis—hypothesis 3

The alternative investment hypothesis states that changes in the stock market are linked to increases in interest in alternative investment opportunities—such as gold and Bitcoin (see Table 7). Thus, in the next step of our research, we analysed whether the number of searches for *gold price* and *bitcoin* can explain the dynamics of stock indices. Again, we estimated GARCH models for the S&P 500 and the Russell 2000 with explanatory variables—but this time we included not only SVIs for stock, but for gold and Bitcoin as well. In the case of both the RUT and S&P 500 returns, in the first period the AR(0)-GARCH(1,1) model was fit, with generalized error distribution (G.E.D.), while in the second period—AR(1)-GARCH(1,1) with Student distribution—for the S&P 500 and AR(1)-ARCH(3) with G.E.D. distribution—for the RUT.

According to our results (see Table 7), the coefficients for the number of *Bitcoin* searches in both sub-samples and both indices are not statistically significant. Yet, the increase in the number of searches for *gold price* is linked to a decrease in S&P 500 prices in both periods. The same conclusion applies to the RUT after the year 2015. In the first period, an increase in *gold price* searches occurs simultaneously with an increase in the volatility of small caps. These results lead to the conclusion that investors consider gold to be a safe-haven asset—not Bitcoin (and this result confirms the findings of e.g. Klein et al. (2018) and Baur et al. (2018)).

5.3 The chicken or the egg—causality analysis (Hypothesis 4)

In the previous parts of the research, we analysed simultaneous changes of investor interest and the dynamics of assets. The results showed that there exist significant relationships between them. As both dependent and explanatory variables were taken from the same day (the closing price of the asset and the total number of searches in a given day), we do not know whether the relationships are causal ones or they are a common reaction to an external event. Therefore, we decided to run yet another test to inspect possible causality patterns. More precisely, we were interested in knowing if increased attention is the cause or effect of the market changes.

In Tables 8, 9 we present the results of the Toda–Yamamoto test for causality (Toda and Yamamoto 1995). We checked the directions of causality between the SVI for a respective keyword and the prices, as well as volumes of given assets. As in the previous two cases, we analysed data in two sub-samples. The numbers in the tables are the p values for the hypothesis of no-causality. Thus, in Table 8 the null hypotheses are as follows: *Prices do not cause investor attention* (upper rows in both panels) and: *Investor interest does not affect prices* (lower rows in both panels). In Table 9 they are: *Volume does not cause investor attention* (upper rows in both panels) and: *Investor interest does not affect the volume* (lower rows in both panels). We are interested in the cases, where the hypothesis was rejected.

In the first sub-period, we observed causality between the number of Google searches and prices in the case of Bitcoin and both stock indices, while the oppo-

Table 7 Estimates of GARCH model the SP500 and RUT returns with SVI for 3 keywords: *bitcoin, stock and gold price*

2013-01-31-2018-01-25		2016-01-04-2018-01-26		
	Coeff.	SD error	t-value	t-prob
SP500 price				
μ	0.001	0.000	3.088	0.002
GSBitcoin(M)	0.001	0.001	1.271	0.204
GSStock(M)	-0.001	0.005	-0.212	0.832
GSGoldPrice(M)	-0.007	0.003	-2.828	0.005
$\omega \cdot 10^6$	7.682	2.723	2.821	0.005
ARCH(α_1)	0.199	0.050	3.938	0.000
GARCH(β_1)	0.682	0.077	8.833	0.000
G.E.D.(DF)	1.556	0.125	12.430	0.000
Russell 2000 price				
2013-01-31-2018-01-25				
μ	0.065	0.038	1.702	0.089
ω	0.086	0.051	1.686	0.092
GSBitcoin(V)	-0.240	0.174	-1.375	0.169
GSStock(V)	2.681	0.477	5.620	0.000
GSGold(V)	0.748	0.307	2.438	0.015
2016-01-04-2018-01-26				
μ	0.09	0.04	2.64	0.01
GSBitcoin(M)	-0.20	0.19	-1.10	0.27
GSStock(M)	0.82	0.71	1.15	0.25
GSGold(M)	-0.61	0.24	-2.60	0.01
AR(1)	-0.07	0.05	-1.52	0.13

Table 7 continued

	2013-01-31-2018-01-25					2016-01-04-2018-01-26				
	Coeff.	SD error	t-value	t-prob		Coeff.	SD error	t-value	t-prob	
ARCH(α_1)	0.059	0.027	2.180	0.030	ω	0.60	0.08	7.04	0.00	
GARCH(β_1)	0.853	0.070	12.150	0.000	GSStock(V)	1.22	0.20	6.09	0.00	
G.E.D.(DF)	1.795	0.147	12.240	0.000	ARCH(α_1)	0.12	0.07	1.89	0.06	
					ARCH(α_2)	0.08	0.05	1.72	0.09	
					ARCH(α_3)	0.14	0.07	2.07	0.04	
					G.E.D.(DF)	1.34	0.14	9.50	0.00	

In the table, we present the estimates of the AR-GARCH model for S&P 500 returns (upper panel) and the Russell 2000 (lower panel). AR(1) denotes the respective AR coefficient α_1 of a conditional mean specification (see the “Appendix” for details). GSStock(M) denotes the number of Google searches of the keyword stock, GSGold(M)—of the keyword *stock*, GSGold(M)—of keyword *gold price*, GSBitcoin—of keyword *bitcoin*—included in a conditional mean equation, while GSStock(V), GSGold(V), GSBitcoin(V)—in a conditional variance equation. ARCH and GARCH denote respectively α and β coefficients from GARCH specification of conditional variance. In the case of both RUT and S&P 500 returns, in the first period the AR(0)-GARCH(1,1) model was fit with generalized error distribution (G.E.D.), while in the second period—AR(1)-GARCH(1,1) with Student distribution—for the S&P 500 and AR(1)-ARCH(3) with G.E.D. distribution—for the RUT

Table 8 *P* values of causality test between investors' interest (SVI) and prices

	BTC	GOLD	SP500	RUT
Period 2013–2015				
<i>Price</i> → <i>SVI</i>	< 0.001	0.001	0.905	0.862
<i>SVI</i> → <i>Price</i>	0.000	0.559	< 0.001	0.036
Period 2016–2018				
<i>Price</i> → <i>SVI</i>	1.000	0.533	0.623	0.832
<i>SVI</i> → <i>Price</i>	< 0.001	0.065	0.870	0.210

In the table we present the *p* values of the null hypothesis of the Toda–Yamamoto test of no causality between investor interest and asset prices in the period 2013–15 (upper panel) and 2016–18 (lower panel). The null hypotheses are as follows: *Prices do not cause investors' attention* (upper rows in both panels) and: *Investors' interest do not cause the prices* (lower rows in both panels). We analyse daily prices of Bitcoin (BTC), gold, the S&P 500 and the Russell 2000 (RUT)

Table 9 *P* values of causality test between the investors' interest (SVI) and volume (VOL)

	BTC	SP500	RUT
Period 2013–2015			
<i>VOL</i> → <i>SVI</i>	0.063	1.000	< 0.001
<i>SVI</i> → <i>VOL</i>	< 0.001	< 0.001	< 0.001
Period 2016–2018			
<i>VOL</i> → <i>SVI</i>	1.000	1.000	1.000
<i>SVI</i> → <i>VOL</i>	< 0.001	< 0.001	< 0.001

In the table we present the *p* values of the null hypothesis of the Toda–Yamamoto test of no causality between investor interest and the trade volume of different assets in the period 2013–15 (upper panel) and 2016–18 (lower panel). The null hypotheses are as follows: *Volume does not cause investors' attention* (upper rows in both panels) and: *Investor interest does not affect volume* (lower rows in both panels). We analyse daily trade volume of Bitcoin (BTC), the S&P 500 and the Russell 2000 (RUT)

site relationship was found in the case of gold and Bitcoin. It suggests that an increase in attention caused changes of prices in the case of the stock indices, while in the case of gold it was price movements that led to changes of attention. Feedback relationships were confirmed when Bitcoin was taken into account. The results obtained for stock indices and gold confirm most of the results received by Vozlyublennia (2014), although our research shows that attention is likely to impact only gold and Bitcoin dynamics—not the stock indices. The results obtained for Bitcoin prices are also in line with the findings of Kristoufek (2015) or Suessmuth (2019).

In the second sub-period, causality between prices and attention was not confirmed in the case of Bitcoin or gold. Causality between attention and stock prices was also not observed. What remained, was the causality relationship between attention to Bitcoin and its prices.

These conclusions have not yet been confirmed by other authors, most probably due to the fact that the period of the most up-to-date data have not been covered in similar research (Aalborg et al. (2019) analysed relationships between SVIs and Bitcoin returns, but did not analyse data in subsamples). We believe that the results could have been caused by a change of investor approaches to information retrieval. The main users of search engines are individual investors who cannot influence the market, even if they react according to attention theory. Perhaps in the first period, after the first research studies showing the importance of Internet searches had been published, institutional investors also started using the Google search engine. This could have contributed to the effect of a self-fulfilling prophecy. In the second sub-period, the interest of institutional investors must have weakened.

Slightly different results have been obtained in the case of the relationships between the volume of assets and investor attention. In the first sub-period, attention caused movement in the volume of Bitcoin and both indices, while the opposite relationship was observed in the case of small caps. In the second sub-period, the latter relationship was not observed, but the test again confirmed causality between attention and volume in the case of the three instruments. We note that causality between SVIs and Bitcoin volume has also been recently confirmed also by Aalborg et al. (2019).

6 Concluding remarks

The aim of the study was to check which financial assets can be considered attention-driven ones and whether an increase in attention can cause movements of asset prices and volume. As a proxy for attention, daily Google searches were used. We analysed the daily prices of two stock indices: the S&P 500 and the RUT (small caps), Bitcoin and gold. As a measure of attention, we took Google search data for the respective keywords (*stock*, *Bitcoin*, *gold prices*). The data were analysed in two sub-periods: 2013–2015 and 2016–2018. We studied the interaction within the same day (using GARCH-type models with explanatory variables) as well as the causality of Granger-type (Toda–Yamamoto test).

The results differed with respect to the analysed assets. In the case of stocks, we confirmed relationships between price volatility and investor attention, while in the case of Bitcoin, relationships were also observed in the conditional mean. Moreover, the volatility of stocks is related to the changing interest of investors in other investment alternatives—namely with gold. As relationships with gold—not with Bitcoin—were confirmed, our study suggests that Bitcoin should not be considered as a safe-haven alternative to stock investing. On the contrary, it is gold (which is not attention-driven) that can be perceived as a safe-haven asset.

Among the analysed assets, Bitcoin is clearly different from the others. Changes of investor attention Granger-caused both the returns and volume of Bitcoin in each period, while in the case of stocks, such relationships were observed only in the first period (2013–2015). On contrary, gold prices did not react to changes of attention, while the attention of investors was Granger-caused by gold price movements (but only over the period 2013–2015). This can indicate that investors in cryptocurrency

markets may belong to a different group than investors in the stock market and that they are more information-driven.

A.1: GARCH models

Let us denote by y_t a vector of time series, and by Ω_{t-1} —the set of information available up to the moment $(t - 1)$. Let us denote the conditional mean of y_t as:

$$y_t = E(y_t | \Omega_{t-1}) + \epsilon_t, \tag{1}$$

where $E(\cdot | \cdot)$ denotes the conditional expectation operator, while ϵ_t is the disturbance term with $E(\epsilon_t) = 0, E(\epsilon_t, \epsilon_s) = 0, \forall t \neq s$. One of the most widely useful specification of the conditional mean is the autoregressive: AR(k) model of the form:

$$y_t - \mu = \sum_{i=1}^k a_i (y_{t-i} - \mu) + \sum_{i=1}^l \delta_i x_{t,i} + \epsilon_t \tag{2}$$

where μ denotes the unconditional mean $x_{i,t}$ denotes the value of the i -th explanatory variable at time t .

If the financial series exhibit significant autocorrelation between observations widely separated in time, we say that the process has a long memory and need to be modelled by a fractionally integrated ARMA process, called ARFIMA (see: Granger (1980) and Granger and Joyeux (1980)). The ARFIMA(n, d, s) model is given by:

$$\Psi(L)(1 - L)^d (y_t - \mu_t) = \Theta(L)\epsilon_t, \tag{3}$$

where $\mu_t = \mu + \sum_{i=1}^l \delta_i x_{t,i}$.

The operator $(1 - L)^d$ accounts for the long memory of the process and is defined as:

$$(1 - L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(d + 1)}{\Gamma(k + 1)\Gamma(d - k + 1)} L^k = 1 - \sum_{k=1}^{\infty} c_k(d)L_k, \tag{4}$$

with $0 < d < 1, c_1(d) = d, c_2(d) = \frac{1}{2}d(1 - d), \dots$ and $\Gamma(\cdot)$ is the Gamma function.

In the case of most financial time-series data the ϵ_t can be decomposed in the following way:

$$\epsilon_t = z_t \sigma_t, \tag{5}$$

where $z_t \sim iid$ with mean 0 and unit variance. If σ_t can be expressed as:

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \tag{6}$$

we say that it follows a GARCH(p, q) process (see: Bollerslev (1986) for details).

Since the pioneering work of Bollerslev Bollerslev (1986) a lot of different specification of the model has appeared. First of all, different distribution of errors has been allowed (e.g. Student, skewed Student or generalized error distribution: GED). From

a variety of different functional specifications of GARCH models, in our study we used a HYGARCH one of Davidson Davidson (2004) to model gold price dynamics in the period (2013–2015). In the model:

$$\sigma_t^2 = \omega[1 - \beta(L)]^{-1} + \{1 - [1 - \beta(L)]^{-1} \cdot \phi(L) \cdot [1 + \alpha(1 - L)^d]\} \epsilon_t^2. \quad (7)$$

We refer the Reader to the original article of Davidson (2004) for more details about the model specification and estimation.

References

- Aalborg HA, Molnár P, de Vries JE (2019) What can explain the price, volatility and trading volume of bitcoin? *Finance Res Lett* 29:255–265. <https://doi.org/10.1016/j.frl.2018.08.010>
- Abraham J, Higdon D, Nelson J, Ibarra J (2018) Cryptocurrency price prediction using tweet volumes and sentiment analysis. *SMU Data Sci Rev* 1(3):1–21
- Alfarano S, Barreda-Tarrazona I, Camacho-Cuena E (2006) On the role of heterogeneous and imperfect information in a laboratory financial market. *CEJOR* 14(4):417–433
- Bai J, Perron P (2003) Computation and analysis of multiple structural change models. *J Appl Econom* 18(1):1–22
- Bank M, Larch M, Peter G (2011) Google search volume and its influence on liquidity and returns of german stocks. *Financ Mark Portfolio Manag* 25(3):239
- Barber BM, Odean T (2007) All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors. *Rev Financ Stud* 21(2):785–818
- Barber BM, Odean T (2013) The behavior of individual investors. In: Constantinides GM, Harris M, Stulz RM (eds) *Handbook of the economics of finance*, vol 2. Elsevier, Amsterdam, pp 1533–1570
- Barclay MJ, Warner JB (1993) Stealth trading and volatility: which trades move prices? *J Finance Econ* 34(3):281–305
- Baur D, Dimpfl T, Kuck K (2018) Bitcoin, gold and the us dollar—a replication and extension. *Finance Res Lett* 25:103–110
- Bedowska-Sojka B (2014) Intraday stealth trading. Evidence from the warsaw stock exchange. *Poznan Univ Econ Rev* 14(1):5–19
- Bijl L, Kringhaug G, Molnár P, Sandvik E (2016) Google searches and stock returns. *Int Rev Financ Anal* 45:150–156
- Black F (1976) Studies of stock market volatility changes. In: *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics Section*, pp 177–181
- Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. *J Comput Sci* 2(1):1–8
- Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. *J Econom* 31:307–327
- Bouri E, Azzi G, Dyrhberg A (2017) On the return-volatility relationship in the bitcoin market around the price crash of 2013. *Econ E-J* 11:1–16
- Caporale GM, Plastun A (2019) The day of the week effect in the cryptocurrency market. *Finance Res Lett* 31:258–269. <https://doi.org/10.1016/j.frl.2018.11.0124>
- Chiah M, Zhong A (2019) Day-of-the-week effect in anomaly returns: international evidence. *Econ Lett* 182(C):90–92
- Da Z, Engelberg J, Gao P (2011) In search of attention. *J Finance* 66(5):1461–1499. <https://doi.org/10.1111/j.1540-6261.2011.01679.x>
- Davidson J (2004) Moment and memory properties of linear conditional heteroscedasticity models, and a new model. *J Bus Econ Stat* 22:16–29
- Dimpfl T, Jank S (2016) Can internet search queries help to predict stock market volatility? *Eur Financ Manag* 22(2):171–192
- Dyrhberg A (2015) Bitcoin, gold and the dollar—a garch volatility analysis. *Finance Res Lett* 16:85–92
- Fama EF (1998) Market efficiency, long-term returns, and behavioral finance. *J Financ Econ* 49(3):283–306
- Granger CWJ (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37:424–438

- Granger CWJ (1980) Long memory relationships and the aggregation of dynamic models. *J Econom* 14:227–238
- Granger CWJ, Joyeux R (1980) An introduction to long-memory time series models and fractional differencing. *J Time Ser Anal* 1:15–29
- Green L, Sung MC, Ma T, Johnson J (2019) To what extent can new web-based technology improve forecasts? assessing the economic value of information derived from virtual globes and its rate of diffusion in a financial market. *Eur J Oper Res* 278(1):226–239. <https://doi.org/10.1016/j.ejor.2019.04.014>
- Gurgul H, Majdosz P (2007) The informational content of insider trading disclosures: empirical results for the polish stock market. *CEJOR* 15(1):1–19. <https://doi.org/10.1007/s10100-006-0016-5>
- Hornung P, Leopold-Wildburger U, Mestel R, Palan S (2015) Insider behavior under different market structures: experimental evidence on trading patterns, manipulation, and profitability. *CEJOR* 23(2):357–373. <https://doi.org/10.1007/s10100-014-0359-2>
- Huber J, Kirchler M, Sutter M (2008) Is more information always better? Experimental financial markets with cumulative information. *J Econ Behav Organ* 65(1):86–104. <https://doi.org/10.1016/j.jebo.2005.05.012>
- Jaffe JF (1974) Special information and insider trading. *J Bus* 47(3):410–428
- Joseph K, Wintoki MB, Zhang Z (2011) Forecasting abnormal stock returns and trading volume using investor sentiment: evidence from online search. *Int J Forecast* 27(4):1116–1127
- Kahneman D (1973) Attention and effort, vol 1063. Citeseer
- Keef SP, Khaled M, Zhu H (2009) The dynamics of the monday effect in international stock indices. *Int Rev Financ Anal* 18(3):125–133
- Klein T, Pham Thu H, Walther T (2018) Bitcoin is not the new gold—a comparison of volatility, correlation, and portfolio performance. *Int Rev Financ Anal* 59(C):105–116. <https://doi.org/10.1016/j.irfa.2018.06.00>
- Kliber A, Marszałek P, Musiałkowska I, Świerczyńska K (2019) Bitcoin: safe haven, hedge or diversifier? Perception of bitcoin in the context of a country's economic situation—a stochastic volatility approach. *Physica A* 524:246–257. <https://doi.org/10.1016/j.physa.2019.04.145>
- Kristoufek L (2015) What are the main drivers of the bitcoin price? evidence from wavelet coherence analysis. *PLoS ONE* 10:1–15. <https://doi.org/10.1371/journal.pone.0123923>
- Li X, Ma J, Wang S, Zhang X (2015) How does google search affect trader positions and crude oil prices? *Econ Model* 49:162–171
- Libby R, Bloomfield R, Nelson MW (2002) Experimental research in financial accounting. *Acc Organ Soc* 27(8):775–810
- Marsden JR, Tung YA (1997) Asymmetric information a laboratory experimental analysis. *Eur J Oper Res* 99(2):256–266. [https://doi.org/10.1016/S0377-2217\(95\)00331-2](https://doi.org/10.1016/S0377-2217(95)00331-2)
- Moat HS, Curme C, Avakian A, Kenett DY, Stanley HE, Preis T (2013) Quantifying wikipedia usage patterns before stock market moves. *Sci Rep* 3:1801
- Nguyen CP, Schinckus C, Nguyen TVH (2019) Google search and stock returns in emerging markets. *Borsa Istanbul Rev*. <https://doi.org/10.1016/j.bir.2019.07.001>
- Preis T, Moat HS, Stanley HE (2013) Quantifying trading behavior in financial markets using google trends. *Sci Rep* 3:1684
- Rutkowska A, Kliber A (2018) Can google trends affect the sentiment of individual investors? The case of the united states. *Math Econ* 14(21):51–70
- Seyhun H (1986) Insiders' profits, costs of trading, and market efficiency. *J Financ Econ* 16(2):189–212. [https://doi.org/10.1016/0304-405X\(86\)90060-7](https://doi.org/10.1016/0304-405X(86)90060-7)
- Shahzad SJH, Bouri E, Roubaud D, Kristoufek L, Lucey B (2019) Is bitcoin a better safe-haven investment than gold and commodities? *Int Rev Financ Anal* 63:322–330. <https://doi.org/10.1016/j.irfa.2019.01.002>
- Shiller RJ (2003) From efficient markets theory to behavioral finance. *J Econ Perspect* 17(1):83–104
- Suessmuth B (2019) Bitcoin and web search query dynamics: is the price driving the hype or is the hype driving the price? CESifo Working Paper Series 7675
- Symitsi E, Chalvatzis KJ (2018) Return, volatility and shock spillovers of bitcoin with energy and technology companies. *Econ Lett* 170(C):127–130
- Tantaopas P, Padungsaksawasdi C, Treepongkaruna S (2016) Attention effect via internet search intensity in asia-pacific stock markets. *Pac Basin Finance J* 38:107–124

- Tetlock PC (2007) Giving content to investor sentiment: the role of media in the stock market. *J Finance* 62(3):1139–1168
- Toda HY, Yamamoto T (1995) Statistical inference in vector autoregressions with possibly integrated processes. *J Econom* 66(1–2):225–250
- Vozlyublennaia N (2014) Investor attention, index performance, and return predictability. *J Bank Finance* 41:17–35

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.