GRA 19703

Master Thesis

How Does Investor Attention Impact the Price of Bitcoin?

Navn: Nicoline Skaug, Minda Marie Bratlie

Start: 15.01.2019 09.00

Finish: 01.07.2019 12.00

How Does Investor Attention Impact the Price of Bitcoin?

A Master Thesis Report for the Degree of

MSc Business with Major in Finance

Authors:

Minda Marie Bratlie and Nicoline Engstrøm Skaug

Supervisor:

Professor Dagfinn Rime

BI Norwegian Business School

July 1, 2019

Oslo, Norway

How Does Investor Attention Impact the Price of Bitcoin?

By Minda Marie Bratlie and Nicoline Engstrøm Skaug

Abstract

Bitcoin has emerged to become the most popular cryptocurrency and its presence has the potential to disrupt existing payment and monetary systems. Over the past decade, the bitcoin price has exhibited extreme volatility, puzzling for both academics and market practitioners. We examine the dynamic relationship between investor attention and the bitcoin price using principal component analysis and vector error correction models and discover that investor attention is an important contributor in bitcoin price formation. Variance decomposition analysis suggests that investor attention explain a significant amount of future variations in the bitcoin price, and investor attention can be used to predict direction of future price change. Our study offers insight into the bitcoin market and the economic impact of investor attention.

Acknowledgements

We would first like to thank our supervisor Professor Dagfinn Rime of the Department of Finance at BI Norwegian Business School. He consistently allowed this paper to be our own work, but guided us in the right direction whenever he saw needed. We would also like to thank our fellow students at BI Norwegian Business School, Gabriel le Gall and Øyvind Emanuel Fryjordet Vennerød, for valuable econometric discussions and programming assistance. Last, but not least, we would like to thank our friends and family for their unconditional love and support during the process of writing this thesis. We are forever grateful.

Contents

List of Tables

List of Figures

1   Research Problem and Motivation

In October 2008, a research paper entitled "Bitcoin: A Peer-to-Peer Electronic Cash System" was published under the pseudonym Satoshi Nakamoto (Nakamoto, 2008).[1] The paper describes an electronic payment system based on cryptographic proof instead of trust with a unit called bitcoin.[2] Bitcoin is the first peer-to-peer network that allows for the proof and transfer of ownership without the need for a trusted third party to verify transactions (Chuen & Deng, 2017).

Since the first bitcoin was created in January 2009 (Wallace, 2011), the popularity and attention around bitcoin and cryptocurrencies has increased, both among private investors and in the corporate world. The New York Stock Exchange created a bitcoin index (New York Stock Exchange, 2018), well-known retailers such as Microsoft, Overstock, and Newegg accept bitcoin as payment (Moreau, 2019), and several start-ups use initial coin offering as a way of raising capital (Nakamura, 2018). Over the past decade, Bitcoin has become a thriving fintech innovation, challenging existing payment and monetary systems. Even at its current early stage, bitcoin and other digital currencies provide a variety of insights about market design and the behavior of buyers and sellers (Böhme, Christin, Edelman, & Moore, 2015).

With the increasing popularity of bitcoin comes the vexing question: What factors determine its monetary value? During 2017, the price of one unit of bitcoin went from \$972 to \$19,214, an increase of more than 1,800%, and the market capitalization of bitcoin went from \$15.9 billion to an all-time-high of \$225.9 billion.[3] The price of one unit of bitcoin is \$8,789, as of May 29, 2019. As bitcoin has no intrinsic value to speak of (Kristoufek, 2013; Mai, Shan, Bai, Wang, & Chiang, 2018; Yermack, 2015), and due to Bitcoin's decentralized structure and primarily online presence, bitcoin derives its value from the value people assign to it (Mai, Bai, Shan, Wang, & Chiang, 2015).[4]

We study if and how investor attention impact the price of bitcoin. Identifying and understanding the factors that drive the price of bitcoin is important with both

---

[1] The number of inventors and identity of Satoshi Nakamoto remains unknown (Bernard, 2018).

[2] We will follow the convention in the computer science literature of using capital-B Bitcoin to refer to the network, and lower-b bitcoin to refer to the unit of account.

[3] https://charts.bitcoin.com/btc/

[4] Note that researchers fail to agree about the intrinsic value of one unit of bitcoin. Some scholars argue that the intrinsic value of one bitcoin equals at least the cost involved in its production through mining (Garcia, Tessone, Mavrodiev, & Perony, 2014).

1

theoretical and practical implications. Theoretical in the sense that investors need predictors to estimate future price fluctuations and calculate the expected return on their investment. Practical in the sense that if bitcoin were to become a means of payment, businesses must understand the volatility of bitcoin before adopting the cryptocurrency.

According to the semi-strong form of asset pricing theory, new information changes expectations of investors and it is instantaneously incorporated into prices when published (Fama, 1970). This assumption requires investors to allocate sufficient attention to the asset (Da, Engelberg, & Gao, 2011a); however, attention is a scarce cognitive resource and investors have limited attention (Kahneman, 1973). When it comes to investment decisions, given the vast amount of information available and the possibility of limited attention, investors ought to be selective in information processing (Peng & Xiong, 2006).

In modern society, the Internet has fundamentally changed how information is distributed and shared. Information on the Internet is updated quickly and spread virally at an exceptional speed, providing first-hand information to investors ahead of other sources (Luo, Zhang, & Duan, 2013). As a result, internet searches have become an important channel through which investors express their demand for public information (Drake, Roulstone, & Thornock, 2012). Researchers argue that social media captures the "wisdom of the crowd," and that social media platforms can reveal information that is relevant to consumer decisions and unobtainable from traditional media (Luo et al., 2013, p. 146).

Due to the Internet's role as a valuable source of information, and given Bitcoin's widespread online presence, it is relevant to investigate how internet search queries impact the price formation of bitcoin. Identifying a feasible link between investor attention and bitcoin price can offer investors, regulators and businesses a new indicator of the future value of the digital currency.

To investigate if and how investor attention impact the price of bitcoin, we obtain data from bitcoin trading markets and internet search queries provided by Google Trends and Wikipedia. We also collect the number of new members on an internet forum (Bitcointalk.org). These variables are used as direct proxies of investors' interest and attention.[5] We employ a Vector Error Correction Model (VECM) to empirically test the relationship between bitcoin value and investor attention. In

---

[5]The difference between indirect and direct proxies is discussed further in Section 3.

addition to sharing many of the benefits of the traditional Vector Autoregression (VAR) model, VECM accounts for endogeneity, autocorrelation, and reverse causality. Furthermore, VECM allows us to capture the interdependencies across time series (Brooks, 2014).

Overall, we discover that investor attention is an important contributor in the price formation of bitcoin. Higher investor attention drives up the price of bitcoin, leading to higher bitcoin trading volume, which again pushes up the bitcoin price. Variance decomposition analysis suggests that investor attention explain a significant amount of future variations in the bitcoin price. Furthermore, investor attention is able to predict future change in bitcoin price with a 51.42% accuracy in a ten-month test period.

Our research makes three main contributions: First, where previous research has focused on time periods (3 months to 3 years) prior to 2016 (Garcia et al., 2014; Kristoufek, 2013, 2015; Mai et al., 2015, 2018), our research focuses on a different and longer time period – July 1, 2015, to December 4, 2018 – capturing both the extreme price increase in 2017 and the "burst of the bubble" throughout 2018. Second, previous papers utilize weekly Google Trends Search Volume Index (SVI) on the search term bitcoin (Kristoufek, 2013; Mai et al., 2015, 2018). We construct a daily Google Trends SVI with the intuition that higher frequency data can capture effects that lower frequency data cannot. Finally, we incorporate a Principal Component Analysis (PCA) index for investor attention instead of including all three investor attention metrics. PCA is a method used to reduce the dimensionality of a large data set by transforming the set of variables into a smaller one that still contains most of the information. The intuition is that this PCA index may have stronger explanatory power on bitcoin price.

The paper is organized as follows: In Section 2, we describe the Bitcoin system and classify bitcoin as an asset. In Section 3, we review previous literature. Section 4 presents the data used in our analysis. Section 5 elaborate on the empirical methodology applied. In Section 6, we present the conducted analyses and overall findings. Finally, Section 7 discusses the implications and limitations of our investigation, and provides suggestions for future studies.

2 The Bitcoin Scheme

The Bitcoin network came into existence January 3, 2009, when Nakamoto mined the genesis block of bitcoin, Block #0, which had a reward of 50 bitcoins (Wallace, 2011). As of May 29, 2019, 578,361 blocks have been created, with a total of 17,729,512 bitcoins in existence.[6] Before analyzing factors that impact the price of bitcoin, it is important to possess an understanding of the technology and core features of the cryptocurrency.[7]

2.1 Technical Description

Bitcoin is the most popular example that uses blockchain technology (Crosby, Nachiappan, Pattanayak, Verma, & Kalyanaraman, 2016). At its core, Bitcoin is a digital public ledger used to administer private property rights of the virtual unit bitcoin. Instead of storing transactions on a single or set of servers, the transaction ledger of Bitcoin is distributed across a network of participating computers (Böhme et al., 2015). Anyone can create a Bitcoin account and get access to the entire database and the complete history of the Bitcoin network (Crosby et al., 2016). The transaction history of the network is stored in a chain of transactions, frequently referred to as the blockchain (Nakamoto, 2008). The blockchain represents all verified and valid transactions between the users in the network. Since blockchain is intrinsically linked to Bitcoin (Crosby et al., 2016), we can explain the concept of blockchain by explaining how Bitcoin works.

Traditionally, financial institutions serve as trusted third parties that validate, safeguard, and preserve any electronic transaction (Crosby et al., 2016). The Bitcoin network, however, lacks a central authority or third party to distribute coins and track who holds which coins (Böhme et al., 2015). Instead, the network is governed by cryptographic rules enforced by transparent computer codes in a decentralized manner (Crosby et al., 2016).

Bitcoins are recorded as transactions. Users do not simply hold bitcoins, in-

---

[6]May 29, 2019, at 14:45 p.m., retrieved from https://www.buybitcoinworldwide.com/how-many -bitcoins-are-there/

[7]The purpose of this section is to give a brief introduction to the technology that Bitcoin is built upon. For a more comprehensive explanation of bitcoin and the Bitcoin network, see Ramzan (2013).

stead they participate in a publicly verifiable transaction network that show all past transactions of users (Böhme et al., 2015). The receiver in a bitcoin transaction verify that the sender can complete the payment by looking at the senders transaction history, ensuring sufficient account balance, before finalizing the transaction. All transactions must be verified before being recorded in the public ledger. The verification process begins by announcing the transaction to all users in the Bitcoin network. As transactions are not announced in the order which they are generated, there is a need for a system to ensure that double-spending does not occur, namely blockchain technology (Crosby et al., 2016).

The process used to secure and verify bitcoin transactions or payments from one user to another in the network is called mining. Mining involves adding bitcoin transaction data to the Bitcoin public ledger (Bitcoin.com, 2017). After announcing a new transaction, transactions occuring at the same time are grouped together into a block of transactions. To ensure that no unauthorized transactions are inserted into the new block, it contains pre-existing contents from the previously issued block (Böhme et al., 2015). By building blocks on top of each other, the blocks are appropriately linear and chronologically ordered, forming the ledger of past transactions, which is the blockchain (Bitcoin.org, 2019). The blockchain structure is what enables any Bitcoin user to verify that a prior transaction did occur and to separate between legitimate transactions and attempts of double-spending (Böhme et al., 2015).

The primary purpose of mining is to enable Bitcoin users to achieve a secure, tamper-resistant consensus (Bitcoin.org, 2019). Bitcoin mining is designed to be difficult and resource intensive so that the number of blocks created each day remains steady over time (Bitcoin.com, 2017). Therefore, Bitcoin has introduced a mathematical puzzle or algorithm that must be solved before the block is accepted by the blockchain (Crosby et al., 2016). The algorithm is based on the pre-existing contents from the previous block and can only be solved using computationally rigorous method. The users donating their computing resources to generate solutions to the algorithm, and thereby generating blocks, are called miners (Böhme et al., 2015). After solving the algorithm, the miner publishes a block containing a proof-of-work that a solution was found, along with all the transactions in the block. After other users verify the published solution, the miners begin working on a new block containing new outstanding transactions (Crosby et al., 2016). In this way, the min-

ing process recursively ensures that the historical ordering of all blocks, is agreed upon by the entire network (Böhme et al., 2015).

Another purpose of the mining process is the creation of new bitcoins. Miners are motivated to assist in the mining process as the miner (or miners) solving the algorithm first is rewarded transaction fees and a subsidy of newly created coins, called block rewards. Mining therefore both serves the purpose of disseminating new coins in a decentralized manner and motivating users to provide security for the system (Bitcoin.com, 2017). Faster computing capabilities are more likely to solve a given algorithm, but speed alone will not guarantee success. As the number of miners in the network changes, the problem difficulty is adjusted to ensure that bitcoins are created at a predetermined rate of roughly 10 minutes (Böhme et al., 2015).

In modern economies, the scarcity of digital money is preserved by legal rules that ensure correct bookkeeping records, as well as by central banks with the power to adjust the quantity of money in circulation. Bitcoin on the other hand, can be understood as the first widely adopted mechanism that provides absolute scarcity of money supply (Böhme et al., 2015). Bitcoins are created at a predictable and decreasing rate. The number of new bitcoins created every year is automatically halved over time until the bitcoin issuance stops completely at a total of 21 million bitcoins in existence. At the point when there are no more bitcoins to mine, Bitcoin miners are likely to be supported exclusively by numerous small transaction fees (Bitcoin.com, 2017).

2.2 Bitcoin – Asset or Currency?

An important question to address is whether to consider bitcoin as a currency or as an asset. There lacks global coordination among authorities regarding how to classify and regulate bitcoin (Bloomberg News, 2018), and the discussion about whether bitcoin is primarily an alternative currency or just a speculative asset is ongoing (European Central Bank, 2012). Selgin (2015) argue that bitcoin has similar features to both commodity and fiat money. He labels bitcoin as a synthetic commodity money due to its similarities with commodity money. However, the majority of scholars suggest that bitcoin should be classified as a financial investment instrument like stock, rather than a currency (Böhme et al., 2015; Glaser, Zimmermann,

Haferkorn, Weber, & Siering, 2014; Kristoufek, 2013; Yermack, 2015).

Money is typically defined by economists as having three attributes – a medium of exchange, a unit of account, and a store of value. Yermack (2015) finds that bitcoin faces challenges in fulfilling all three criteria. Bitcoin faces challenges as a medium of exchange due to the difficulty of the procurement of new bitcoins. In terms of being a useful unit of account, a crucial problem arises from the relatively high cost of one bitcoin compared to most products and services. Finally, bitcoin faces challenges as a store of value due to its high volatility. Instead of behaving like a currency according to the criteria widely used by economists, bitcoin appears to behave more like a speculative investment.

Böhme et al. (2015) argue that Bitcoin's design presents distinctive risks that differ from other payment methods and stores of value: market risk, the shallow market problem, counterparty risk, transaction risk, operational risk, privacy-related risk, and legal and regulatory risk.[8] In particular, when comparing the coefficient of variation for the daily USD/BTC exchange rate with other currency exchanges, the researchers find that bitcoin is 41 times more volatile than the USD/EUR exchange rate.

Fred Ersham, Co-founder of Coinbase, a digital exchange where merchants and consumers can transact with bitcoin, estimate that in 2014, 80% of the activity on the site was related to speculation (Nathan, 2014). Further, Glaser et al. (2014) find strong indications that users who are particularly uninformed, are not primarily interested in an alternative transaction system when approaching digital currencies, but rather seek to participate in an alternative investment vehicle. These findings support Mr. Ersham's estimates. The researchers conclude that most users (by volume) treat their bitcoin investments as speculative assets rather than as means of payments.

Even though authorities, central banks, and researchers have yet to come to a conclusive agreement on how to classify bitcoin, its market characteristics suggest that we can study bitcoin using models for stocks (Mai et al., 2018). Thus, for the remainder of this paper, we consider bitcoin as an asset, and our work is based on theory connecting investor attention and asset value.

---

[8]For further and more elaborate explanation of these risks, see Böhme et al. (2015).

3 Review of Literature

The efficient market hypothesis is one of the most established theories in finance literature (Fama, 1970; Samuelson, 1965). The theory states that all relevant information may change expectations of investors and affect the stock price. The impact of the release of a piece of information depends on the current market expectation, which has been formed on the basis of previous information. Without the release of new information, no price movement will occur (Fama, 1991). Several studies provide a theoretical framework in which limited attention can affect asset pricing statics and dynamics (Hirshleifer & Teoh, 2003; Merton, 1987; Peng & Xiong, 2006; Sims, 2003).

3.1 Literature with Indirect Proxies for Investor Attention

Investors face a daunting choice when looking for a security to buy. Odean (1999) proposes that investors manage the problem of choosing among thousands of possible stock purchases by limiting their search to stocks that have recently caught their attention. Therefore, when testing theories of investor attention, researchers face a considerable challenge: there is no direct measure for investor attention. To overcome this, researchers rather use indirect proxies for investor attention (Da et al., 2011a).

Barber and Odean (2008) use the following three proxies in their study about the effect of attention on the buying behavior of investors: (i) a stock's abnormal daily trading volume; (ii) the stock's (previous) one-day return; and (iii) whether the firm appeared in that day's news. They find that investors are net buyers of what they call "attention-grabbing stocks," and conclude that an increase in individual investor attention results in temporary positive price pressure.

Trading volume is also used as a proxy for investor attention in the study of Gervais, Kaniel, and Mingelgrin (2001) who investigate the future evolution of stock prices after extreme trading activity. Gervais et al. find that stocks experiencing unusually high (low) trading volume over a day or a week tend to appreciate (depreciate) over the course of the following month.

Gervais et al. (2001) and Lou (2014) use firm advertising expenses as a proxy for investor attention, and discover that firms with greater advertising expenditures, ceteris paribus, have a larger number of both individual and institutional investors as

well as better liquidity of their common stock (Gervais et al., 2001). Furthermore, increased advertisement spending is associated with a contemporaneous rise in retail buying and abnormal stock returns (Lou, 2014). Seasholes and Wu (2004) show that, on price limit days, a higher percentage of purchases are made by first-time buyers compared to non-price limit days (Seasholes & Wu, 2004). Their interpretation of this behavior is that individual investors, particularly first-time buyers, are attracted by the stock hitting a price limit. Therefore, individuals become net buyers of stocks that catch their attention.

Da et al. (2011a) argue that proxies of indirect attention, such as those mentioned above, make the critical assumption that investors paid attention to a stock if its name was mentioned in the news media, or that its return or turnover was extreme. However, return or turnover can be driven by other factors unrelated to investor attention, and "a news article in the *Wall Street Journal* does not guarantee attention unless investors actually read it" (Da et al., 2011a, p. 1462). The researchers suggest using aggregate search frequency from Google as a direct measure of investor attention.

3.2   Literature with Direct Proxies for Investor Attention

The Google Trends SVI measures the extent to which a particular term or phrase is searched for using the Google search engine. Google searches are reported as an index over time for a particular search term, either globally or in specific regions. Each index is defined from 0 to 100, where 100 represents the point in time where the search term had the largest share of the total queries in the specific region. All other values are relative to this maximum (Google, 2019).

There are primarily three reasons for utilizing Google Trends SVI as a measure of attention. First, internet users commonly use a search engine to collect information, and Google is ranked number one by market share: In 2018, Google had a search engine market share of 73.62% (Net Market Share, 2019). The search volume reported by Google is therefore likely to represent the internet search behavior of the population as a whole (Da et al., 2011a). Second, search queries are revealed attention measures: if investors search for information regarding bitcoin on Google, they are likely paying attention to it. Therefore, aggregate search frequency in Google is a direct and appropriate measure of attention (Da et al., 2011a). Third,

there is strong empirical evidence that demonstrates the predictive power of Google search and how it can be used to forecast anything from tourism, automobile sales, home sales, and influenza epidemics (Choi & Varian, 2012; Ginsberg et al., 2009).

Da et al. (2011a) were the first to propose using search frequency in Google as a direct measure of investor attention. They investigate the attention paid towards particular stocks (Russel 3000 stocks) by examining the Google Trends SVI for respective stock ticker symbols. They discover that time series correlations between Google Trends SVI (log) and alternative weekly measures of attention, such as extreme returns, turnover, and news, are on average positive, but that the level of correlation is low. Furthermore, they find strong evidence that Google Trends SVI captures the attention of both individual and retail investors.

Following a different approach, Da, Engelberg, and Gao (2011b) show that Google Trends SVI predicts returns around earnings announcements. These findings suggest that "search volume contains value-relevant information that is not incorporated into prices until the announcement" (Da et al., 2011b, p. 2). More recently, Da, Engelberg, and Gao (2014) construct a Financial and Economic Attitudes Revealed by Search (FEARS) index by aggregating the volume of queries related to household concerns (e.g., recession, unemployment, and bankruptcy) as a new measure of investor sentiment. They discover that the FEARS index is able to predict short-term return reversals, temporary increases in volatility, and flows out of equity funds and into bond funds.

Choi, along with Google's Chief Economist Hal Varian, find that Google search queries have the potential to describe short-term interest in various of economic activities in real time (Choi & Varian, 2012). Similarly, if investors search more on pre-event days, the changes in pre-announcement price and volume reflect more information, and there is less response in price and volume when earnings are announced (Drake et al., 2012).

Regulators conclude that social media is "landscape-shifting" and that the use of social media by the financial services industry is rapidly accelerating (The Securities and Exchange Commission, 2012). As information is easily and almost freely available in today's digital world, researchers have started using other online metrics as direct proxies for investor attention, such as internet message board activity (Antweiler & Frank, 2004; Tumarkin & Whitelaw, 2001), customer reviews (Luo & Zhang, 2013), search queries on Wikipedia (Kristoufek, 2013; Moat et al., 2013),

text content on Twitter (Bollen, Mao, & Zeng, 2011; Mai et al., 2018), and social media (Chen, De, Hu, & Hwang, 2014; Luo & Zhang, 2013; Luo et al., 2013; Mai et al., 2018).

An early study that examine financial internet blogs and message boards find that higher activity in financial blogs and message boards have a statistically significant influence on stock returns: A positive shock to message board postings predicts negative stock returns on the next day (Antweiler & Frank, 2004).

Luo and Zhang (2013) investigate the predictive power of both consumer buzz and web traffic on firm value. Their findings indicate that buzz and traffic explain a substantial portion of the total variance of firm value, which confirm the relevance of consumer word-of-mouth and website visits on firm value. In a similar way, Chen et al. (2014) discover that opinions revealed in posts on one of the largest investment-related social media websites in the United States strongly predict future stock returns and earnings surprises. By contrast, Tumarkin and Whitelaw (2001) find that message board activity can predict neither stock returns nor volume. Instead, they discover the opposite: market information influences message board activity. Overall, there are inconsistencies across previous studies regarding the relationship between investor attention and financial markets.

3.3 Literature on Bitcoin and Investor Attention

As we classify bitcoin as an asset, there are prominent differences between bitcoin and stocks that should be highlighted. For example, bitcoin has no discounted future cash flows (e.g., dividends) and therefore no intrinsic value to speak of (Kristoufek, 2013; Mai et al., 2018; Yermack, 2015). The bitcoin market also has limited depth, which indicates that it is costly to trade (Mai et al., 2018). Together, these features indicate that previous research should be used with caution when investigating the connection between social media information and bitcoin value. However, as argued by Mai et al. (2018), bitcoin exhibits several unique features that suggest a significant predictive relationship between social media metrics and the value of bitcoin.

Kristoufek (2013) studies the dynamic relationship between bitcoin price and two proxies for investors' interest and attention: search queries on Google and search frequency on Wikipedia. He discovers a strong positive correlation between

bitcoin price and Google Trends SVI (weekly) and Wikipedia (daily): 87.86% and 82.71%, respectively. He also finds that when the price is high (above trend), an increase in interest pushes the price further atop. Converse, if the price is below trend, growing interest drives the price down.

Following a different approach, Kristoufek (2015) investigates the main drivers of the bitcoin price and discovers that the price is driven by investors' interest in the cryptocurrency. The relationship appears to be most evident in the long-run, while it is prone to bubbles and busts in the short-run. During periods of explosive prices, interest drives the prices further up, and during rapid declines, interest pushes the price further down.

Mai et al. (2018) extend previous literature and investigate whether and to what extent social media impacts the value of bitcoin. They conduct a sentiment analysis of messages on an internet forum (Bitcointalk.org) and Twitter. Overall, they find that social media is an important predictor of the future value of bitcoin. Specifically, "more bullish (or bearish) forum posts are significantly associated with higher (or lower) next-day bitcoin market price" (Mai et al., 2018, p. 22).

3.4 Literature on Bubbles

The considerable fluctuations in bitcoin price are not suggestive of a constant fundamental value, but rather due to a substantial speculative component. A possible specification of this speculative component is the existence of a bubble in the bitcoin market (Cheah & Fry, 2015). Bubbles are typically associated with dramatic asset price increase followed by a collapse. This situation can occur if an investor holds an asset because they believe they can sell it at a higher price than some other investor even though the asset's price exceeds its fundamental value (Brunnermeier, 2008).

Brunnermeier (2008) suggests that bubbles can emerge when investors have heterogeneous beliefs and face short-sale constraints. Investors have heterogeneous beliefs if they start with a different prior belief distribution than others, possibly due to psychological biases (Brunnermeier, 2008). The combination of heterogeneous beliefs and short-sale constraints can result in overpricing as optimists push up the asset price, while pessimists cannot counterbalance this due to short-sale constraints (Miller, 1977). Heterogeneous belief bubbles are categorized by large trading vol-

ume and high price volatility (Scheinkman & Xiong, 2003), both of which can be observed in the bitcoin market.

Garcia et al. (2014) find that when bitcoin investors make investment decisions they tend to partake in self-reinforcing feedback loops when discussing, searching for, and utilizing information in social media and news reporting. They provide evidence for inter-individual influence where increased interest encourages individual investors to purchase bitcoins, which drives up the price, and eventually feeds back on search volume, fueling a feedback loop.

As bitcoin and other cryptocurrencies are relatively new, there exists only a limited amount of research, especially when it comes to investor attention and bitcoin price formation. Two prominent contributions are Kristoufek (2013) and Mai et al. (2018), which we build on and extend in three ways: First, our research is done over a different and longer time period – July 1, 2015, to December 4, 2018 – capturing both the extreme price increase in 2017 and the "burst of the bubble" throughout 2018. Second, previous papers (Kristoufek, 2013; Mai et al., 2015, 2018) utilize weekly Google Trends SVI on the search term bitcoin. We construct a daily Google Trends SVI with the intuition that higher frequency data can capture effects that lower frequency data cannot. Finally, we incorporate an investor attention PCA index that may have stronger explanatory power on bitcoin price than the individual investor attention metrics separately.

4 Data

4.1 Measures for Market Price of Bitcoin

The focal point of our empirical analysis is the market price of bitcoin. We study the dynamic relationship between the natural logarithm of price and other variables. A convenient feature of *ln* (price) is that the continuously compounded return in bitcoin is the first difference of *ln* (price):

$$r_t = ln\left(\frac{P_t}{P_{t-1}}\right) = ln(P_t) - ln(P_{t-1}) \tag{1}$$

where r_t denotes the continuously compounded return at time t, P_t is the bitcoin market price at the end of day t, and ln denotes the natural logarithm. Changes in log price is widely used in asset pricing research (Campbell, 1996).

Unlike stock markets, there are no official bitcoin exchanges, but instead, there are crypto exchanges around the world that operate 24/7. Therefore, there is no official bitcoin price. Our data set is comprised of daily market price, USD/BTC exchange rate, from BitStamp Ltd., the top bitcoin exchange by volume at time of data collection.[9]

In addition to price, we include volatility of bitcoin return and trading volume to control for other observable variations in the bitcoin market. The daily trading volume is the amount of bitcoin traded for USD in a 24-hour period. The daily trading volume was obtained from Coinmarketcap.com, denoted V_t.[10] Figure 1 displays the time series of the bitcoin price and the bitcoin trading volume.

To capture the effects on bitcoin price brought about by uncertainty, we include a risk measure of bitcoin price using the volatility of bitcoin returns. To measure the volatility of return, we apply the Exponentially Weighted Moving Average (EWMA) model. EWMA is essentially an extension of the historical average volatility measure where more recent observations have a stronger impact on the forecast of volatility than older data points (Brooks, 2014). EWMA tracks changes in volatility with the formula:

$$\sigma_t^2 = (1 - \lambda) \sum_{j=0}^{\infty} \lambda^j (r_{t-j} - \bar{r})^2 \qquad (2)$$

where σ_t^2 is the estimate of volatility on day t, $\bar{r}$ is the average return estimated over the observations, and λ is the multiplier that determines how much weight is assigned to the most recent observation. By setting $\bar{r}$ equal to zero, and $\lambda = 0.94$, as suggested by RiskMetrics (Brooks, 2014, p. 421), the formula can be expressed as:

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1 - \lambda) r_{t-1}^2 \qquad (3)$$

where σ_t^2 is the estimate of volatility on day t and r_{t-1} is the most recent daily percentage price. Observations on bitcoin price in the period between January 1 and June 30, 2015, was used to train the EMWA formula, before applying it on the time period under investigation.

---

[9]All data was collected during February 2019.

[10]The bitcoin market has two measures of volume: trading volume and transaction volume. Trading volume refers to the exchange of bitcoin for fiat currency (off-chain), whereas transaction volume is the amount of bitcoin used for transactions (on-chain). As we investigate investor attention, we assume trading volume to be the most relevant measure of investor activity in the bitcoin market.
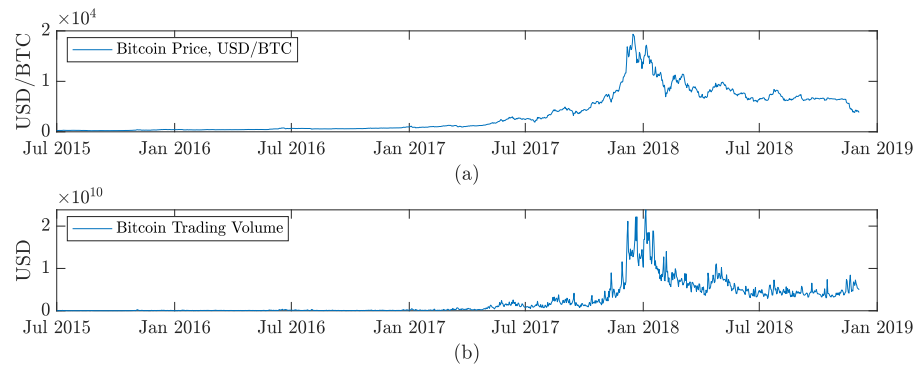
Figure 1: Time Series Plot of Bitcoin Price and Bitcoin Trading Volume

Note. The daily time series of bitcoin price and bitcoin trading volume are shown in subfigure (a) and (b), respectively. The sample period is July 1, 2015, to December 4, 2018. Source: USD/BTC exchange rate, obtained from BitStamp Ltd., and trading volume, obtained from Coinmarketcap.com.

4.2   Direct Proxies for Investor Attention

4.2.1   Google Trends SVI

For time frames shorter than six months, Google Trends provide daily data, whereas for longer time frames, Google Trends provide weekly data. As we study a time frame of four years, we are not able to obtain daily Google Trend SVI on the search term bitcoin directly. Instead, we construct a daily Google Trends SVI that is similar to the weekly Google Trends SVI with the intuition that higher frequency data can capture effects that lower frequency data cannot. We download daily Google Trends SVI data in pairs of two days, where the last observation in the first period overlaps with the first observation in the next period. The data is then chained and scaled to obtain one value for each day.[11]

Furthermore, we utilize the opportunity to limit searches related to finance by using the category filter option. This allows us to obtain searches where the user has entered a finance-related website after searching for the term bitcoin.[12] Neither Kristoufek (2013) nor Mai et al. (2015, 2018) implement this option. However, as we are interested in studying the impact of investor attention on bitcoin price, this is precisely the group of people whose attention we desire to capture (Choi & Varian, 2012; Da et al., 2011a).

---

[11]See Appendix A for further procedure information.

[12]Google treat search queries as case-insensitive and all letters in a search phrase are interpreted in lower case. Therefore, searches for bitcoin, Bitcoin and BITCOIN will return the same result, hence the Google Trends SVI is not affected by capitalized letters.
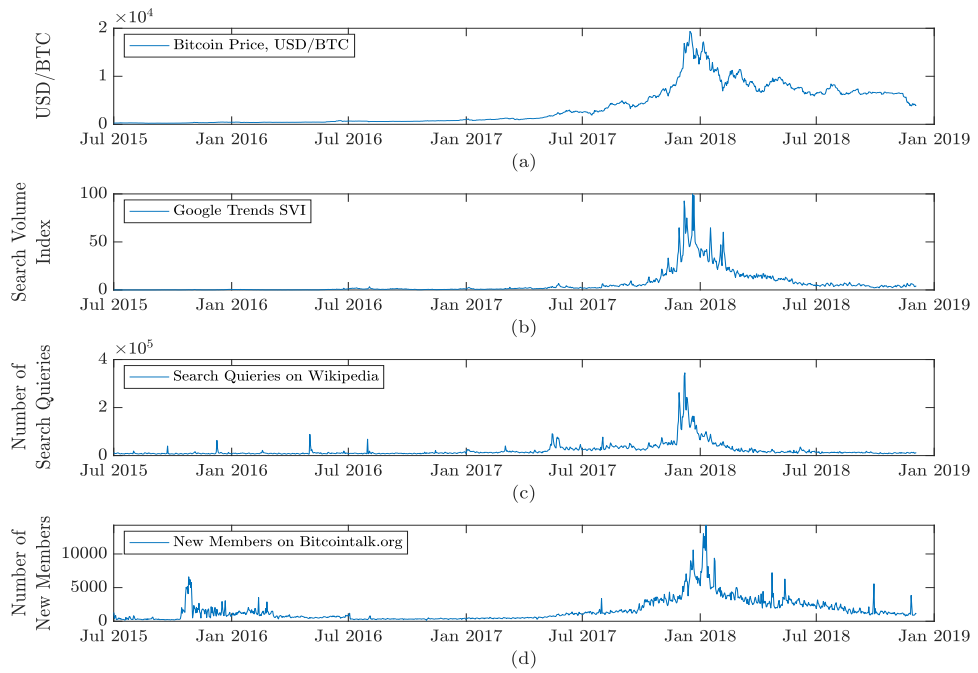
Figure 2: Time Series Plot of Bitcoin Price and Investor Attention Metrics

Note. The daily time series of bitcoin price, Google Trends SVI, search queries on Wikipedia, and new members on Bitcointalk.org are shown in subfigures (a), (b), (c), and (d), respectively. Although we do not make statistical inferences from these figures, we note that the time series of the three investor attention metrics peak at approximately the same time as when the bitcoin price reaches its highest of $19,345 on December 16, 2017. The sample period is July 1, 2015, to December 4, 2018. Source: USD/BTC exchange rate, obtained from BitStamp Ltd., the search volume index on the search term bitcoin, obtained from Google Trends (see Appendix A), the number of search queries for the search term bitcoin, obtained from Wikipedia, and number of new members, obtained from Bitcointalk.org.

4.2.2 Wikipedia

We follow Kristoufek (2013) and collect daily search queries volume on Wikipedia for the search term bitcoin. According to Kristoufek, the frequency of searches related to a digital currency is a good measure of potential investors interest in the currency. There are two measures of Wikipedia user activity: the number of page views per day and the number of page edits that take place at a specific Wikipedia page per week (Moat et al., 2013). We focus on the number of page views per day as we assume this measure to be the best at capturing investor attention. However, note that this variable measures both investors' and the general public's interest in bitcoin, which makes it difficult to determine if the search for information was used to guide an investment decision or not.

4.2.3 Bitcointalk.org

We obtain daily statistics of new members from the dominant community platform Bitcointalk.org (Smyth, 2013). To the best of our knowledge, the only researchers to have used this variable when investigating bitcoin price formation are Ciaian, Rajcaniova, and Kancs (2016). The scholars argue that new members on Bitcointalk.org captures both the size of the Bitcoin economy and the attention-driven behaviour of new bitcoin investors. Furthermore, messages on Bitcointalk.org are proved to have a strong impact on future bitcoin value (Mai et al., 2018). Therefore, we suspect the number of new members per day on Bitcointalk.org to have a stronger explanatory power on the price of bitcoin than general search indices.

Figure 2 displays the time series of the bitcoin price and the investor attention metrics. In addition to a striking similarity between the time series, we find high positive correlations between the bitcoin price and the investor attention variables: 80.96%, 60.53%, and 80.91%, for Google Trends SVI, search queries on Wikipedia, and new members on Bitcointalk.org, respectively.

4.2.4 Principal Component Analysis

The three investor attention variables exhibit high positive correlations: 83.96% between Google Trends SVI and search queries on Wikipedia, 77.54% between Google Trends SVI and new members on Bitcointalk.org, and 56.71% between new members on Bitcointalk.org and search queries on Wikipedia.[13] As the three variables are all intended to measure the impact of investor attention on the bitcoin price, and given their high positive correlation, we can use a PCA index to reduce the dimensionality of the VECM.

PCA is a statistical procedure that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of linearly uncorrelated variables called principal components (Brooks, 2014). The main idea of PCA is to reduce the dimensionality of a data set that consists of several significantly correlated variables, as it allows for the retention of the most important variation and influence present in the variables. The first principal component accounts for 78.61% of the variance in the investor attention variables. Figure 3 displays the time series of the investor attention PCA index and the investor attention metrics.

---

[13]The correlation was calculated before log transformation.
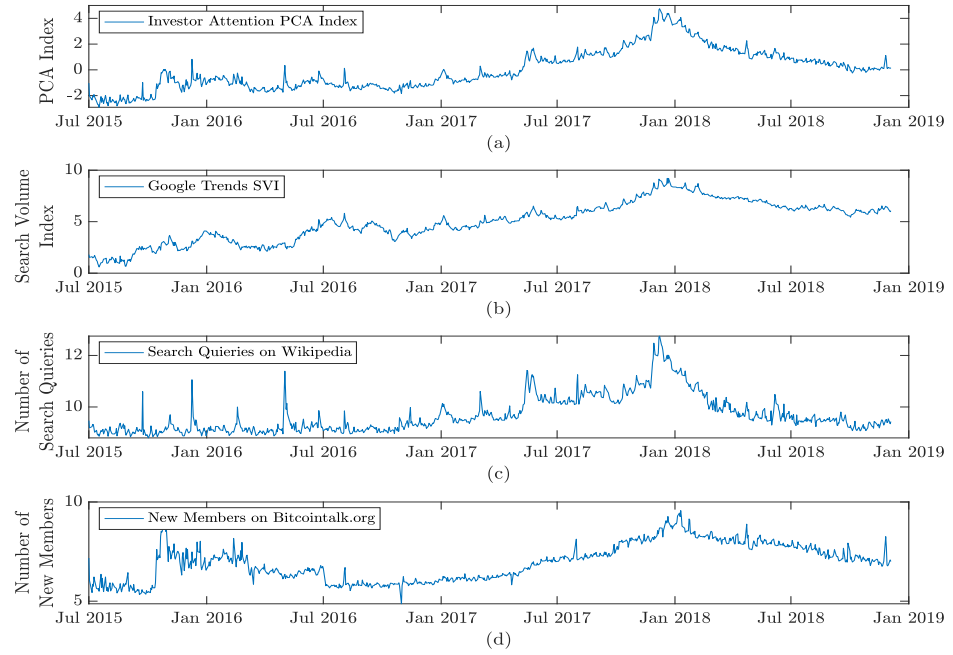
Figure 3: Time Series Plot of Investor Attention PCA Index and Investor Attention Metrics

Note. The daily time series of the investor attention PCA index, Google Trends SVI (log), search queries on Wikipedia (log), and new members on Bitcointalk.org (log) are shown in subfigures (a), (b), (c), and (d), respectively. The investor attention PCA index accounts for 78.61% of the variance in the three investor attention variables. The sample period is July 1, 2015, to December 4, 2018. Source: The investor attention PCA index is calculated based on the three investor attention variables. The search volume index on the search term bitcoin, obtained from Google Trends (see Appendix A), the number of search queries for the search term bitcoin, obtained from Wikipedia, and number of new members, obtained from Bitcointalk.org.

4.3   Control Variables

We include two traditional control variables from the financial market: Standard & Poor's (S&P) 500 index ($sp500$) and stock market volatility (vix), both obtained from Yahoo Finance.[14] As the bitcoin market is a 24/7 market, we obtain more observations of bitcoin data than from the financial market. We therefore exclude observations on non-trading days, which results in a sample size of $n = 863$. Table 1 summarizes the key measures and statistics.

---

[14] Some economists compare bitcoin to gold (Baur, Dimpfl, & Kuck, 2018; Dyhrberg, 2016), and researchers use gold as a control variable when investigating the price formation of bitcoin (Kristoufek, 2015; Mai et al., 2015, 2018). The COMEX gold price ($gold$), obtained from Investing.com, was initially included in the VECM. However, as we found no statistically significant relationship between the bitcoin price and $gold$, see Table B1 in Appendix B, the COMEX gold price was excluded to reduce dimensionality.

Table 1: Key Measures and Summary Statistics

Variable Definition	Symbol	Mean	Std	Median	Min	Max	Correlation[a]
Bitcoin Market Variables							
Bitcoin price (log)	$ln(P)$	7.394	1.319	7.070	5.354	9.851	1.000
Volatility of bitcoin returns[b]	σ^2	0.151	0.135	0.113	0.008	0.650	0.611
Daily trading volume (log)	V	20.005	2.099	19.603	16.664	23.895	0.930
Investor Attention Metrics							
Google Trends SVI for bitcoin (log)[c]	$googletrends$	4.986	1.917	5.139	0.596	9.210	0.810
Wikipedia search queries for bitcoin (log)	$wikipedia$	9.690	0.721	9.462	8.800	12.750	0.605
Daily new members on Bitcointalk.org (log)	$newmembers$	6.953	0.926	7.007	4.868	9.566	0.809
Control Variables							
S&P 500 closing price (log)	$sp500$	7.763	0.127	7.767	7.512	7.983	0.831
COBE volatility index	vix	14.844	4.573	13.430	9.140	40.740	−0.160

Note. This table presents summary statistics for all daily variables. Std is the standard deviation, Min is the minimum value, and Max is the maximum value. The sample period is July 1, 2015, to December 4, 2018. Days with missing observations (i.e. non-trading days) were removed, resulting in a sample size of $n = 863$. Source: Authors' calculations and values obtained from MATLAB.

[a]The Pearson's correlation coefficients are calculated between USD/BTC and the model variables before log transformation.

[b]The volatility of bitcoin returns, σ^2, was scaled by multiplying each data point with 100 to avoid small coefficients.

[c]The Google Trends SVI, $googletrends$, was scaled by multiplying each data point with 100 prior to the log transformation to avoid negative coefficients.

19

5   Empirical Methodology

5.1   Model Specifications

To study the dynamic relationship between the bitcoin price and investor attention, we use a VECM to capture the interdependencies across the different time series. VECM is a generalization of the VAR system that incorporates the long-term corrections so that both short-term and long-term dynamics can be studied when cointegration is present (Brooks, 2014).[15]

There are particularly four reasons why we use a VECM instead of a traditional multiple regression model (cf. Antweiler and Frank, (2004)). First, VECM works as an extension of the VAR model as it allows us to model the recursive relationship between interdependent variables. This indicates that there is not need to separate or specify variables as endogenous and exogenous, instead we can treat all variables as jointly endogenous. Second, we do not need prior knowledge about the factors that influence a variable as there can be numerous cointegrating relationships in a VECM. Third, the model allows for both autocorrelation and cross-correlation, which indicates that we can better understand the dynamic relationships among the variables. Finally, by using Granger causality we test if past values of investor attention variables are useful for predicting the bitcoin market variables, and to establish the causality between the variables.

We examine a model that includes variables on daily observations of bitcoin market activities: price ($ln(P)$), volatility (σ^2), and trading volume (V). The model also includes an investor attention PCA index (IA) which is calculated using Google Trends SVI ($googletrends$), search queries on Wikipedia ($wikipedia$), and the number of new members on Bitcointalk.org ($newmemebers$). We also include the control variables discussed in Section 4.3.

We now present how we determine the appropriate model. See Appendix C for more details on the performed model specification tests. First, we test the variables for unit roots and determine if they are stationary. Traditional regression techniques, including VAR, encounter problems when applied to non-stationary data. Regressing two independent random walk processes can yield a significant coefficient estimate and a high R^2 even if the variables are not related. We use both

---

[15]Unless otherwise indicated, all subsequent theory in this section is based on Brooks (2014).

GRA 19703

the Augmented Dickey-Fuller (ADF) unit root test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) stationarity test. By using both tests, we can compare the results to see if we obtain the same conclusion. Such joint use of stationarity and unit root tests is known as confirmatory data analysis.

According to the ADF test, σ^2 and vix are stationary, and the remaining model variables are non-stationary with one order of integration. The KPSS test concludes that all variables are non-stationary with one order of integration. Due to conflicting results, we test for autocorrelation and find that all time series exhibit autocorrelation up to lag 20, see Figure C1. After differencing the autocorrelated time series once, none appear to exhibit autocorrelation, see Figure C2. Therefore, we favor the KPSS test results that suggests all time series to be non-stationary with one order of integration. We use the Akaike Information Criterion (AIC) to choose the optimal lag length k in the model, which is standard in econometrics literature (Mai et al., 2018). We estimate VAR models with lag length varying from 0 to 12 and compute the AIC. The optimal lag length prove to be $k = 3$.

Furthermore, we use the Johansen multiple trace test to test for cointegration rank. The trace test is a sequential, joint hypothesis testing procedure where the null hypothesis is that the number of cointegrating vectors is less than or equal to rank (r) against an unspecified or general alternative that there are more cointegrating vectors than r. The test is repeated until the first null hypothesis is not rejected. We discover three cointegration relationships, $r = 3$, which indicates that there are stable and long-term equilibrium relationships among the variables. On the premise of the existence of cointegration relationships, a VECM can be further conducted.

5.2 VECM

It is possible to model the relationship between non-stationary variables using VAR by taking the first difference of each time series, but this approach can suffer from misspecification biases if cointegration is present (Lütkepohl, 2005). A more sophisticated approach is to utilize the VECM as this model yields more efficient estimators of cointegrating time series. The VECM uses a vector of error correction terms that is equal in length to the number of cointegrating relationships added to the relationship. Formally, a VECM with p variables, k lags, and cointegration

21

order r has the following form:

$$\Delta \boldsymbol{Y}_t = \sum_{j=1}^{k-1} \Gamma_j \Delta \boldsymbol{Y}_{t-k} + \alpha \beta^T \boldsymbol{Y}_{t-k} + \boldsymbol{\mu} + \epsilon_t \tag{4}$$

where Δ is the first difference operator, $\boldsymbol{Y}_t$ is a $p \times 1$ vector with order of integration 1, $\boldsymbol{\mu}$ is a $p \times 1$ vector that represents the linear trend, k is the lag length, and ϵ is the residual vector. In addition, Γ_j is a $p \times p$ matrix that indicates the short-term relationships among variables, β is a $p \times r$ matrix that represents the long-term relationship between the cointegrating vectors, and α is a $p \times r$ matrix denoting the speed of which the variables adjust to the long-term equilibrium. If variables are cointegrated, there will be stationary linear combinations of $\beta^T \boldsymbol{Y}_t$ (i.e. a long-term equilibrium relation), although $\boldsymbol{Y}_t$ itself is non-stationary.

The difference between a VECM and a VAR model with first-difference variables is the additional $\beta^T \boldsymbol{Y}_{t-1}$, known as the error correction term. As such, the VECM is a special case of the general VAR system expressed as an equivalent VAR:

$$\boldsymbol{Y}_t = (I_k + \alpha \beta^T + \Gamma_1) \boldsymbol{Y}_{t-1} + \sum_{j=2}^{k-1} (\Gamma_j - \Gamma_{j-1}) \boldsymbol{Y}_{t-j} + \Gamma_{k-1} \boldsymbol{Y}_{t-k} + \boldsymbol{\mu} + \epsilon_t \tag{5}$$

where I_k is a $k \times k$ identity matrix.

6 Analyses and Results

To test how investor attention impacts the price of bitcoin, we examine the effect of an investor attention PCA index on the bitcoin price using a VECM. The model includes daily measures of the bitcoin market variables ($ln(P)$, σ^2, and V), the investor attention PCA index (IA), as well as the control variables discussed in Section 4.3. We use a model with lag length $k = 3$, according to the AIC, and rank $r = 3$, according to the Johansen trace test.

6.1 VECM Analyses

The VECM estimation output consists of two parts. The first part reports the long-term relationship between the cointegrating vectors in the model. In the presence of more than one cointegrating vector, Johansen and Juselius (1990) consider the first cointegrating vector to be the most useful. Therefore, as we investigate the impact on the bitcoin price, we normalize the variable $ln(P)$ in the first cointegrating vector

to express it as a function of the remaining model variables. This cointegrating vector is given by:

$$\beta_1^T = \begin{bmatrix} 1.000 & 2.687 & -0.064 & -0.634 & -4.221 & 0.012 & 26.050 \end{bmatrix} \quad (6)$$

These values represent the coefficients for $ln(P)$ (normalized to one), σ^2, V, IA, $sp500$, vix, and a constant term. The cointegrating vector β_1^T can be expressed as:

$$ln(P) = -2.687\sigma^2 + 0.064V + 0.634IA + 4.221sp500 - 0.012vix - 26.050 \quad (7)$$

From Equation 7, it can be seen that, ceteris paribus, each percentage point increase in bitcoin trading volume will cause an increase of 0.064% in the bitcoin price, and each percentage point increase in investor attention will cause an increase of 0.634%.

The second part of the VECM estimation output reports the short-term relationship among the model variables and the estimated error correction terms. The error correction terms are the speed of adjustment for each variable in every cointegrating relationship, and they indicate how fast the variables converge to its long-term equilibrium value. Table 2 depicts both the estimated error correction terms (Coint.Eq.1-3) and the short-term characteristics of the bitcoin market.

Days with higher bitcoin price tend to precede days with higher bitcoin trading volume. Volatility of bitcoin return is statistically significant on both the bitcoin price and the trading volume, with a negative impact on the first and a positive impact on the latter. Days with higher bitcoin trading volume tend to predate days with higher bitcoin price, lower volatility, and lower trading volume. Furthermore, days with both higher bitcoin price and higher trading volume tend to precede days of lower investor attention.

Investor attention exhibit a strong statistically significant relationship with bitcoin price and trading volume: Days with higher investor attention tend to predate days with both higher bitcoin price and higher trading volume. A one percent increase in investor attention is associated with an increase in the bitcoin price by 0.58% and a 0.09% increase in trading volume. Conversely, days with higher investor attention tend to occur before days with lower investor attention.

As discussed in Section 3.4, Garcia et al. (2014) claim that bitcoin investors tend to partake in self-reinforcing feedback loops when making investment decisions as increased interest encourages individual investors to purchase bitcoins, which drives

23

Table 2: VECM Error Correction Estimates with PCA Index

Independent Variables	Dependent Variables			
	$\Delta ln(P)_t$	$\Delta \sigma_t^2$	ΔV_t	ΔIA_t
Coint.Eq.1	0.004***	0.005***	0.033***	−0.010
	(0.002)	(0.001)	(0.009)	(0.008)
Coint.Eq.2	−0.002	0.001	−0.040***	0.009
	(0.002)	(0.001)	(0.009)	(0.008)
Coint.Eq.3	−0.002	0.003***	−0.030***	−0.005
	(0.002)	(0.001)	(0.009)	(0.008)
$\Delta ln(P)_{t-1}$	−0.009	0.036	0.017**	−0.012*
	(0.001)	(0.063)	(0.007)	(0.007)
$\Delta \sigma_{t-1}^2$ [a]	−0.041**	−0.024	0.024***	−0.006
	(0.018)	(0.033)	(0.004)	(0.004)
ΔV_{t-1}	0.690***	−0.943**	−0.345***	−0.086**
	(0.216)	(0.388)	(0.042)	(0.042)
ΔIA_{t-1}	0.581***	−0.486	0.094**	−0.296***
	(0.193)	(0.347)	(0.038)	(0.038)
$\Delta sp500_{t-1}$	0.010	−0.015	0.001	−0.001
	(0.001)	(0.012)	(0.001)	(0.001)
Δvix_{t-1}	−1.360	1.392	0.007	0.077
	(1.248)	(2.244)	(0.244)	(0.245)

Note. This table presents the estimated error correction terms and first lag coefficients of the short-term relationship between the model variables. $ln(P)$ is the bitcoin price (log), σ^2 is the volatility of bitcoin returns, V is the daily trading volume (log), and IA is the investor attention PCA index. Other variables are as defined in Table 1. For concreteness, the control variables are not displayed among the dependent variables. Lag length $k = 3$. Number of cointegrated relationships $r = 3$. The sample period is July 1, 2015, to December 4, 2018. Days with missing observations (i.e. non-trading days) were removed, resulting in a sample size of $n = 859$. Estimated, asymptotic standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
[a] The volatility of bitcoin returns, σ^2, was scaled by multiplying each data point with 100 to avoid small coefficients.

up the price, and eventually feeds back on search volumes. We find indications of such a feedback loop as higher investor attention pushes up the bitcoin price, leading to higher trading volume, which again drives up the price of bitcoin. However, our result also suggests the presence of possibly stabilizing mechanisms as days with higher bitcoin price, higher trading volume, and higher investor attention tend to precede days with lower investor attention. Furthermore, days with higher trading volume tend to predate days with lower trading volume. Overall, we find evidence that investor attention does impact the price of bitcoin.

To confirm this result, we perform a Granger causality test between $ln(P)$ and lagged values of σ^2, V, and IA. These relationships are individually ($\chi^2 = 18.509$, $p < 0.001$ for σ^2; $\chi^2 = 7.892$, $p = 0.048$ for V; $\chi^2 = 11.654$, $p = 0.009$ for IA) and jointly ($\chi^2 = 12.178$, $p = 0.058$) significant, which indicate that past values of σ^2, V, and IA cause changes in the bitcoin price. Finally, none of the control variables are statistically significant on the bitcoin price.

6.2 Forecast Error Variance Decomposition

Given the estimated effects of investor attention on the bitcoin price, we now proceed to examine a more practical question: To what extent does investor attention explain the future variance of the bitcoin price? To answer this, we derive the FEVD measure. FEVD determines how much of the forecast error variance of each variable can be explained by exogenous shocks to the other variables. Thus, FEVD is comparable to R^2 in regression models and provides insight about the relative importance of each variable (Brooks, 2014). FEVD is defined as:

$$FEVD_{jk}, s = \sum_{i=0}^{s-1} \frac{p_{jk,i}^2}{MSE_k(s)} \tag{8}$$

where $MSE_k(s)$ is the mean squared error of s-step forecast of variable k, and $p_{jk,i}$ is the effect of a one-unit shock to variable j on k given by the impulse response function.

We follow Luo and Zhang (2013) and Mai et al. (2018) and evaluate the FEVD values at 20 days. Table 3 provides a breakdown of the forecast error variance of the bitcoin price that can be attributed to shocks to itself or other variables in the model. As expected, the bitcoin price variable accounts for the largest fraction of its own forecast error variance. The explanatory power for the bitcoin market variables is 94.36% when the investor attention PCA index included, and 98.74% when excluded. The reduction of 4.38% is distributed with 79.22% onto the investor attention PCA index and 20.78% onto the control variables. Overall, the FEVD analysis emphasizes that investor attention adds meaningful explanatory power for the bitcoin price after controlling for investor attention and other control variables.

Table 3: Variance of Bitcoin Price Explained by Different Variables

	VECM Including PCA Index	VECM Excluding PCA Index
Bitcoin Market Variables		
$ln(P)$	91.58	97.75
σ^2	0.80	0.82
V	1.98	0.17
Total	*94.36*	*98.74*
Investor Attention Index		
IA	3.47	–
Total	*3.47*	–
Control Variables		
$sp500$	2.12	1.20
vix	0.05	0.06
Total	*2.17*	*1.26*

Note. This table presents a breakdown of the forecast error variance (FEVD) of bitcoin price explained by shocks to different model variables. A generalized FEVD computation method was applied, which compute variance decompositions using one-standard-deviation innovation shocks. Variables are as defined in Table 1. Numbers are in percentages, evaluated at 20 days. The sample period is July 1, 2015, to December 4, 2018. Days with missing observations (i.e. non-trading days) were removed, resulting in a sample size of $n = 858$.

6.3   Forecast Accuracy

Another practical question to investigate: Does investor attention help forecast future bitcoin price? To answer this, we test the predictive power of investor attention variables by conducting an out-of-sample forecast. Out-of-sample forecasting is regarded useful to test the accuracy of a model (Brooks, 2014). We reserve the last quarter of the period under investigation, from January 30 to December 4, 2018, as the test period (215 days). First, the model is estimated with the observations prior to the test period. It is then re-estimated day-by-day until the last day of the entire sample. The updated parameters are used to generate new one-day ahead forecasts. Such iterative forecasts imitate actual behavior in real time and are frequently used in economics (Meese & Rogoff, 1983).

We measure forecast accuracy using root mean square error (RMSE) and mean absolute error (MAE). The RMSE is defined as $\sqrt{\frac{1}{n}\sum(actual - predicted)^2}$, and the MAE is defined as $\frac{1}{n}\sum|actual - predicted|$, where n is the number of forecasting periods. Smaller RMSE and MAE indicate better model performance.

Table 4: Comparison of Forecasting Accuracy

	Model		
	3-Day Moving Average	Random Walk	VECM
RMSE	0.0036	0.0022	0.0022
MAE	0.0433	0.0325	0.0330

Note. This table presents the forecasting accuracy measures for three different models, which are calculated using a 215-day period from January 30, 2018, to December 4, 2018. For each day, every model is estimated using all the data prior to that day. The model's parameters are used to forecast the bitcoin price the next day. The root mean square error (RMSE) is the standard deviation of the residuals (prediction error), and the mean absolute error (MAE) is the average of all absolute errors. Smaller RMSE and MAE indicate better model performance. Source: Authors' calculations and values obtained from MATLAB.

A three-day moving average model is used as a benchmark for evaluating the accuracy of VECM forecasts. In addition, we include a random walk model. Table 4 presents the results. In our test period, the random walk model has the lowest RMSE and MAE of 0.0022 and 0.0325, respectively. The VECM has the same RMSE (0.0022), but a slightly higher MAE of 0.0330. The three-day moving average model has the highest RMSE and MAE of 0.0036 and 0.0433, respectively. The obtained results indicate that the random walk model is slightly more accurate at predicting future bitcoin price compared to the VECM. Nevertheless, the fact that the VECM outperforms the three-day moving average benchmark provides evidence that investor attention is relevant in the determination of future bitcoin price.

Furthermore, it is important to note that a random walk model does not predict change as it assumes that the best estimate for today's price is yesterday's price. The VECM that includes the investor attention PCA index, on the other hand, is able to predict the correct direction of change in 51.42% of the cases. This indicates that investors should consider investor attention metrics when developing a strategy for investing in bitcoin.

6.4 Robustness Tests

To assess our results, we conduct robustness tests. To investigate if we obtain similar results as those obtained when using the investor attention PCA index, we model a VECM that includes all investor attention metrics, and three separate VECMs that include the metrics individually. The results of the VECM including all three in-

Table 5: VECM Error Correction Estimates All Investor Attention Variables

Independent Variables	Dependent Variables		
	$\Delta ln(P)_t$	$\Delta\sigma_t^2$	ΔV_t
$\Delta ln(P)_{t-1}$	-0.004	0.036	0.015^{**}
	(0.035)	(0.063)	(0.007)
$\Delta\sigma_{t-1}^2$ [a]	-0.041^{**}	-0.047	0.021^{***}
	(0.018)	(0.033)	(0.004)
ΔV_{t-1}	0.679^{***}	-0.997^{**}	-0.355^{***}
	(0.216)	(0.390)	(0.043)
$\Delta googletrends_{t-1}$	0.205	-0.927^{***}	0.043
	(0.146)	(0.264)	(0.029)
$\Delta wikipedia_{t-1}$	0.371^{**}	-0.550^{*}	0.034
	(0.161)	(29.092)	(0.032)
$\Delta newmembers_{t-1}$	0.372^{**}	0.160	0.023
	(0.182)	(32.817)	(0.037)

Note. This table presents the first lag coefficients of the short-term relationship between the model variables. Estimates for controls are not displayed. $ln(P)$ is the bitcoin price (log), σ^2 is the volatility of bitcoin returns, and V is the daily trading volume (log). Other variables are as defined in Table 1. For concreteness, the investor attention metrics are not displayed among the dependent variables. The sample period is July 1, 2015, to December 4, 2018. Days with missing observations (i.e. non-trading days) were removed, resulting in a sample size of $n = 858$. Lag length $k = 3$. Number of cointegrated relationships $r = 3$. Estimated, asymptotic standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
[a]The volatility of bitcoin returns, σ^2, was scaled by multiplying each data point with 100 to avoid small coefficients.

vestor attention metrics are shown in Table 5. The results are largely consistent with the results obtained in Table 2, except that when separating the investor attention variables, none exhibit a statistically significant impact on trading volume.

When including the investor attention variables individually, the results hold for the Google Trends SVI, see Table D1, and for the new members on Bitcointalk.org, see Table D3. However, the results do not hold when including solely the Wikipedia variable, see Table D2. A possible reason for the conflicting result is that, as mentioned in Section 4.2.2, this variable measures both the interest of investors and of the general public, making it difficult to determine if the information search is used to guide an investment decision or not. This indicates that search queries on Wikipedia may not be a useful direct proxy when measuring investor attention.

To ensure that the results in Table 2 are not driven by specific events in a specific

time frame, we introduce an exogenous time dummy variable with the investor attention PCA index and estimate the VECM again. The time dummy variable takes a value of one if it is after January 1, 2017.[16] The results shown in Table D4 are largely consistent with our main findings.

7   Discussion and Conclusion

Bitcoin and other cryptocurrencies provide unique benefits, including lower transaction costs and stimulus for fintech innovation (Brito & Castillo, 2013). The price dynamic of bitcoin has been a controversial topic attracting public attention due to price surges with the potential of abnormal profits in short periods of time. Lack of understanding the price fluctuations of bitcoin can prevent the cryptocurrency from achieving its full potential. We therefore sought to quantify the dynamic relationship between investor attention and the price of bitcoin.

7.1   Implication of Investigation

To the best of our knowledge, this study is the first to investigate the economic impact of investor attention on the bitcoin price in the period between July 2015 and December 2018, when the cyptocurrency experienced a rapid price increase accompanied by a sudden plunge. Our results show that investor attention is an important contributor in the price formation of bitcoin. With increased investor attention comes an increase in the price of bitcoin, which leads to higher bitcoin trading volume and, consequently, a rise in the bitcoin price. Evidence from the Granger causality test and variance decomposition suggests that investor attention does have a predictive power on the bitcoin price. Furthermore, out-of-sample forecasting shows that investor attention is able to predict future change in bitcoin price with a 51.42% accuracy.

Our findings have implications for investors and cryptocurrency adopters. For investors, the predictive relationship between investor attention and bitcoin market variables suggests that investors can use information demand on online channels to obtain novel information about the future bitcoin market. Improved predictability

---

[16]The cut-off point is placed on January 1, 2017, as this is the beginning of the rapid bitcoin price increase towards its peak in December 2017.

of the fluctuations of the cryptocurrency can improve its reliability as a standard component of investment portfolios.

Companies and cryptocurrency adopters should strategically consider their decision to adopt bitcoin as a means of payments. If bitcoin is accepted, companies ought be able to identify and understand the factors that drive the price of bitcoin, as the future value of accounts receivables are affected by price fluctuations. Companies should also be aware of the uncertainties surrounding the implementation of new regulations of the bitcoin market.

7.2   Limitations and Suggestion for Future Studies

Our research has several limitations in its data sources and analysis methods which suggest possible extensions. With regards to the investor attention variables there lacks a possibility to distinguish between unique Internet Protocol (IP) addresses. For instance, each search query on Wikipedia is counted, even if it comes from the same IP address. Thus, the overall number of search queries may not coincide with the amount of individual investors. The Google Trends SVI is limited to finance related searches. Even though we assume this to increase the probability to capture the attention of bitcoin investors, it is difficult to distinguish if the search was done by someone interested in investing or not. Furthermore, our study is limited to observations on trading days, hence observations on non-trading days were removed. This data alteration may have had an impact on the results.

As mentioned earlier, our research is the first to investigate the price formation of bitcoin in this period of interest. As we use solely direct proxies for investor attention, we recommend future studies to investigate the relationship between investor attention and the bitcoin market in a more sophisticated way, through for example sentiment analysis of forum posts and tweets, as done by Mai et al. (2018). Moreover, we examine the association between the USD/BTC exchange rate and investor attention. As other exchange rates surpass USD in bitcoin trading volume, future studies should investigate if the discovered relationships also hold for other exchange rates.

Bibliography

Antweiler, W., & Frank, M. Z. (2004). Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance*, *59*(3), 1259–1294. doi: 10.1111/j.1540-6261.2004.00662.x

Barber, B. M., & Odean, T. (2008). All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *The Review of Financial Studies*, *21*(2), 785–818. doi: 10.1093/rfs/hhm079

Baur, D. G., Dimpfl, T., & Kuck, K. (2018). Bitcoin, Gold and the US Dollar – A Replication and Extension. *Finance Research Letters*, *25*, 103–110. doi: 10.1016/j.frl.2017.10.012

Bernard, Z. (2018, November 10). Everything you need to know about bitcoin, its mysterious origins, and the many alleged identities of its creator. Retrieved from https://www.businessinsider.fr/us/bitcoin-history -cryptocurrency-satoshi-nakamoto-2017-12

Bitcoin.com. (2017). Bitcoin Mining. Retrieved from https://www.bitcoin.com/ bitcoin-mining

Bitcoin.org. (2019, February 4). Frequently Asked Questions – Bitcoin. Retrieved from https://bitcoin.org/en/faq#general

Bloomberg News. (2018, March 26). What the Worlds Governments Are Saying About Cryptocurrencies. Retrieved from https://www.bloomberg.com

Böhme, R., Christin, N., Edelman, B., & Moore, T. (2015). Bitcoin: Economics, Technology, and Governance. *Journal of Economic Perspectives*, *29*(2), 213–38. doi: 10.1257/jep.29.2.213

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter Mood Predicts the Stock Market. *Journal of Computational Science*, *2*(1), 1–8. doi: 10.1016/j.jocs.2010.12 .007

Brito, J., & Castillo, A. (2013). *Bitcoin: A Primer for Policymakers*. Mercatus Center at George Mason University.

Brooks, C. (2014). *Introductory Econometrics for Finance* (3rd ed.). Cambridge University Press.

Brunnermeier, M. K. (2008). Bubbles. *The New Palgrave Dictionary of Economics*, *1–8*(2), 578–583. doi: 10.1007/978-1-349-58802-2

Campbell, J. Y. (1996). Understanding Risk and Return. *Journal of Political*

Economy, *104*(2), 298–345. doi: 10.1086/262026

Cheah, E.-T., & Fry, J. (2015). Speculative Bubbles in Bitcoin Markets? An Empirical Investigation Into the Fundamental Value of Bitcoin. *Economics Letters*, *130*, 32–36. doi: 10.1016/j.econlet.2015.02.029

Chen, H., De, P., Hu, Y. J., & Hwang, B.-H. (2014). Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media. *The Review of Financial Studies*, *27*(5), 1367–1403. doi: 10.1093/rfs/hhu001

Choi, H., & Varian, H. (2012). Predicting the Present with Google Trends. *Economic Record*, *88*, 2–9. doi: 10.1111/j.1475-4932.2012.00809.x

Chuen, D. L. K., & Deng, R. H. (2017). *Handbook of Blockchain, Digital Finance, and Inclusion: Cryptocurrency, FinTech, InsurTech, Regulation, ChinaTech, Mobile Security, and Distributed Ledger*. Academic Press.

Ciaian, P., Rajcaniova, M., & Kancs, d. (2016). The Economics of BitCoin Price Formation. *Applied Economics*, *48*(19), 1799–1815. doi: 10.1080/00036846 .2015.1109038

Crosby, M., Nachiappan, N., Pattanayak, P., Verma, S., & Kalyanaraman, V. (2016). BlockChain Technology: Beyond Bitcoin. *The Applied Innovation Review*, *2*, 1–16. Retrieved from http://scet.berkeley.edu/wp-content/uploads/AIR-2016 -Blockchain.pdf

Da, Z., Engelberg, J., & Gao, P. (2011a). In Search of Attention. *The Journal of Finance*, *66*(5), 1461–1499. doi: 10.1111/j.1540-6261.2011.01679.x

Da, Z., Engelberg, J., & Gao, P. (2011b). In Search of Fundamentals. *AFA 2012 Chicago Meetings Paper*. doi: 10.2139/ssrn.1589805

Da, Z., Engelberg, J., & Gao, P. (2014). The Sum of All FEARS: Investor Sentiment and Asset Prices. *The Review of Financial Studies*, *28*(1), 1–32. doi: 10.1093/ rfs/hhu072

Drake, M. S., Roulstone, D. T., & Thornock, J. R. (2012). Investor Information Demand: Evidence from Google Searches Around Earnings Announcements. *Journal of Accounting Research*, *50*(4), 1001–1040. doi: 10.1111/j.1475-679X.2012.00443.x

Dyhrberg, A. H. (2016). Bitcoin, Gold and the Dollar – A GARCH Volatility Analysis. *Finance Research Letters*, *16*, 85–92. doi: 10.1016/j.frl.2015.10 .008

European Central Bank. (2012). *Virtual Currency Schemes*. Retrieved from https://

www.ecb.europa.eu/pub/pdf/other/virtualcurrencyschemes201210en.pdf

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, *25*(2), 383–417. doi: 10.2307/2325486

Fama, E. F. (1991). Efficient Capital Markets: II. *The Journal of Finance*, *46*(5), 1575–1617. doi: 10.1111/j.1540-6261.1991.tb04636.x

Garcia, D., Tessone, C. J., Mavrodiev, P., & Perony, N. (2014). The Digital Traces of Bubbles: Feedback Cycles Between Socio-Economic Signals in the Bitcoin Economy. *Journal of the Royal Society Interface*, *11*(99), 20140623. doi: 10.1098/rsif.2014.0623

Gervais, S., Kaniel, R., & Mingelgrin, D. H. (2001). The High-Volume Return Premium. *The Journal of Finance*, *56*(3), 877–919. doi: 10.1111/0022-1082 .00349

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting Influenza Epidemics Using Search Engine Query Data. *Nature*, *457*(7232), 1012–1015. doi: 10.1038/nature07634

Glaser, F., Zimmermann, K., Haferkorn, M., Weber, M., & Siering, M. (2014). Bitcoin – Asset or Currency? Revealing Users' Hidden Intentions. *Twenty Second European Conference on Information Systems, Tel Aviv*.

Google. (2019, January 27). How Trends Data is Adjusted. Retrieved from https://support.google.com

Hirshleifer, D., & Teoh, S. H. (2003). Limited Attention, Information Disclosure, and Financial Reporting. *Journal of Accounting and Economics*, *36*(1-3), 337–386. doi: 10.1016/j.jacceco.2003.10.002

Johansen, S., & Juselius, K. (1990). Maximum Likelihood Estimation and Inference on Cointegration – With Applications to the Demand for Money. *Oxford Bulletin of Economics and statistics*, *52*(2), 169–210. doi: 10.1111/j.1468 -0084.1990.mp52002003.x

Kahneman, D. (1973). *Attention and Effort* (Vol. 1063). Prentice-Hall Inc., NJ: Prentice-Hall.

Kristoufek, L. (2013). Bitcoin Meets Google Trends and Wikipedia: Quantifying the Relationship Between Phenomena of the Internet Era. *Scientific Reports*, *3*, 3415. doi: 10.1038/srep03415

Kristoufek, L. (2015). What Are the Main Drivers of the Bitcoin Price? Evidence from Wavelet Coherence Analysis. *PloS one*, *10*(4). doi: 10.1371/journal

.pone.0123923

Lou, D. (2014). Attracting Investor Attention Through Advertising. *The Review of Financial Studies*, *27*(6), 1797–1829. doi: 10.1093/rfs/hhu019

Luo, X., & Zhang, J. (2013). How Do Consumer Buzz and Traffic in Social Media Marketing Predict the Value of the Firm? *Journal of Management Information Systems*, *30*(2), 213–238. doi: 10.2753/MIS0742-1222300208

Luo, X., Zhang, J., & Duan, W. (2013). Social Media and Firm Equity Value. *Information Systems Research*, *24*(1), 146–163. doi: 10.1287/isre.1120.0462

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer Science & Business Media.

Mai, F., Bai, Q., Shan, J., Wang, X. S., & Chiang, R. H. (2015). The Impacts of Social Media on Bitcoin Performance. *Thirty Sixth International Conference on Information Systems*. Retrieved from http://www.fmaconferences.org/ Orlando/Papers/Bitcoin_FMA.pdf

Mai, F., Shan, Z., Bai, Q., Wang, X., & Chiang, R. H. (2018). How Does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis. *Journal of Management Information Systems*, *35*(1), 19–52. doi: 10.1080/07421222.2018.1440774

Meese, R. A., & Rogoff, K. (1983). Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample? *Journal of International Economics*, *14*(1-2), 3–24. doi: 10.1016/0022-1996(83)90017-X

Merton, R. C. (1987). A Simple Model of Capital Market Equilibrium with Incomplete Information. *The Journal of Finance*, *42*(3), 483–510. doi: 10.1111/j.1540-6261.1987.tb04565.x

Miller, E. M. (1977). Risk, Uncertainty, and Divergence of Opinion. *The Journal of Finance*, *32*(4), 1151–1168. doi: 10.1111/j.1540-6261.1977.tb03317.x

Moat, H. S., Curme, C., Avakian, A., Kenett, D. Y., Stanley, H. E., & Preis, T. (2013). Quantifying Wikipedia Usage Patterns Before Stock Market Moves. *Scientific reports*, *3*, 1801. doi: 10.1038/srep01801

Moreau, E. (2019, January 19). 15 Major Retailers and Services That Accept Bitcoin. Retrieved from https://www.lifewire.com

Nakamoto, S. (2008). Bitcoin: A Peer-To-Peer Electronic Cash System. , 1–9. Retrieved from https://bitcoin.org/bitcoin.pdf

Nakamura, Y. (2018). Startups Are Raising Billions Using Initial Coin Offerings.

Retrieved from https://www.bloomberg.com

Nathan, A. (2014). Interview with Fred Ersham. *Goldman Sachs Global Macro Research*, *21*, 8–9. Retrieved from https://www.paymentlawadvisor.com/files/2014/01/GoldmanSachs-Bit-Coin.pdf

Net Market Share. (2019). Search Engine Market Share. Retrieved from https://netmarketshare.com

New York Stock Exchange. (2018). NYSE® Bitcoin Index (NYXBT). , 1–13. Retrieved from https://www.nyse.com/publicdocs/nyse/indices/NYSE_Bitcoin_Index_Methodology.pdf

Odean, T. (1999). Do Investors Trade Too Much? *American Economic Review*, *89*(5), 1279–1298. doi: 10.1257/aer.89.5.1279

Peng, L., & Xiong, W. (2006). Investor Attention, Overconfidence and Category Learning. *Journal of Financial Economics*, *80*(3), 563–602. doi: 10.3386/w11400

Ramzan, Z. (2013). Bitcoin. Retrieved from https://www.khanacademy.org

Samuelson, P. (1965). Proof That Properly Anticipated Prices Fluctuate Randomly. *Industrial Management Review*, *6*(2). doi: 10.1142/9789814566926_0002

Scheinkman, J. A., & Xiong, W. (2003). Overconfidence and Speculative Bubbles. *Journal of Political Economy*, *111*(6), 1183–1220. doi: 10.1086/378531

Seasholes, M., & Wu, G. (2004). Profiting From Predictability: Smart Traders, Daily Price Limits, and Investor Attention. *University of California, Berkeley, Working Paper*.

Selgin, G. (2015). Synthetic Commodity Money. *Journal of Financial Stability*, *17*, 92–99. doi: 10.2139/ssrn.2000118

Sims, C. A. (2003). Implications of Rational Inattention. *Journal of Monetary Economics*, *50*(3), 665–690. doi: 10.1016/S0304-3932(03)00029-1

Smyth, L. (2013, March 13). The Demographics of Bitcoin (Part 1 Updated). Retrieved from https://spacedruiddotcom.wordpress.com/2013/03/04/the-demographics-of-bitcoin-part-1-updated/

The Securities and Exchange Commission. (2012). Investment Adviser Use of Social Media. *National Examination Risk Alert*, *2*(1), 1–7. Retrieved from https://www.sec.gov/about/offices/ocie/riskalert-socialmedia.pdf

Tumarkin, R., & Whitelaw, R. F. (2001). News or Noise? Internet Postings and Stock Prices. *Financial Analysts Journal*, *57*(3), 41–51. doi: 10.2469/faj.v57

.n3.2449

Wallace, B. (2011, November 23). The Rise and Fall of Bitcoin. Retrieved from

https://www.wired.com/2011/11/mf-bitcoin/

Yermack, D. (2015). *Is Bitcoin a Real Currency? An Economic Appraisal*. Elsevier.

doi: 10.3386/w19747

Appendices

A Constructing Daily Google Trends SVI

This section presents procedure information on how the daily Google Trends SVI was constructed. Daily data was downloaded in pairs of two, by applying a Python web-crawler. The pairs comprise of the last observation in the first period, which overlaps with the first observation in the next period (i.e., the first period = 1.7.2015 – 2.7.2015, the second period = 2.7.2015 – 3.7.2015, and so on). Within each pair, one day has the value of 100, and the other a relative value depending on the search level on that specific day. By comparing the pairs, we are able to determine whether the Google Trends SVI increase or decrease from one day to the next. After chaining the pairs, we are left with two observations for each day. To obtain one value the following formula was applied:

$$constructed_t = constructed_{t-1}\left(\frac{googledaily_t}{googledaily_{t-1}}\right) \tag{A1}$$

where $constructed_t$ is the constructed daily Google Trends SVI series at time t, $googledaily_t$ is the first Google Trends SVI value in the current pair, and $googledaily_{t-1}$ is the last Google Trends SVI value in the previous pair.

Figure A1 depicts both the weekly and the constructed daily Google Trends SVI. It is evident that both time series share similar movements and spikes. However, the constructed time series exhibit more rapid fluctuations and higher frequency. We suspect that the constructed time series is able to capture effects and fluctuations in the bitcoin price that the weekly time series does not.
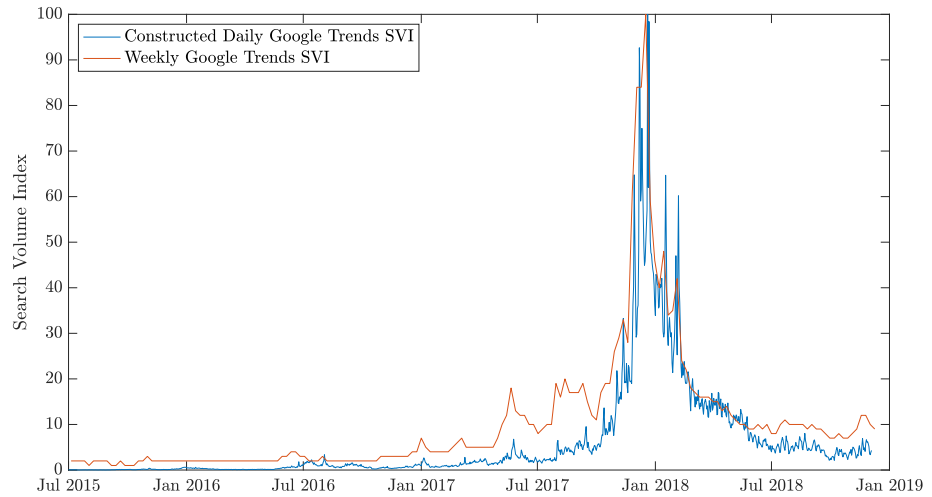
Figure A1: Comparing Daily and Weekly Google Trends SVI

This figure depicts both the weekly Google Trends SVI and the constructed daily Google Trends SVI time series. The sample period is July 5, 2015, to December 9, 2018, for the weekly Google Trends SVI, and July 1, 2015, to December 4, 2018, for the constructed daily Google Trends SVI. It can be seen that the constructed daily Google Trends SVI is approximately comparable to the weekly Google Trends SVI as both time series share similar movements and spikes. However, the constructed Google Trends SVI exhibit more rapid fluctuations and higher frequency. Source: Google Trends and authors' calculations.

B   VECM – Including COMEX Gold Price

Table B1: VECM Error Correction Estimates Including COMEX Gold Price

Independent Variables	Dependent Variables			
	$\Delta ln(P)_t$	$\Delta\sigma^2_t$	ΔV_t	ΔIA_t
$\Delta ln(P)_{t-1}$	0.001	0.001	2.526**	0.017*
	(0.001)	(0.001)	(6.345)	(0.007)
$\Delta\sigma^2_{t-1}$ [a]	-0.001**	-0.001	-1.706***	0.024
	(0.001)	(0.001)	(3.339)	(0.004)
ΔV_{t-1}	0.004***	0.007**	-78.721***	-0.351
	(0.003)	(0.002)	(38.988)	(0.042)
ΔIA_{t-1}	-0.007**	0.005	-33.879**	0.091***
	(0.003)	(0.002)	(34.687)	(0.037)
$\Delta sp500_{t-1}$	0.005	0.001	-148.971	0.004
	(0.009)	(0.001)	(117.297)	(0.126)
$\Delta gold_{t-1}$	-0.001	0.001	-1.676	-0.002
	(0.001)	(0.001)	(1.454)	(0.002)
Δvix_{t-1}	-0.061	-0.013	157.711	0.028
	(0.017)	(0.012)	(225.185)	(0.241)

Note. This table presents the first lag coefficients of the short-term relationship between the model variables. As can be seen, $gold_{t-1}$ is not statistically significant on bitcoin price, thus we omit this variable from the final VECM. $ln(P)$ is the bitcoin price (log), σ^2 is the volatility of bitcoin returns, V is the daily trading volume (log), and IA is the investor attention PCA index. Other variables are as defined in Table 1. For concreteness, the control variables are not displayed among the dependent variables. Lag length $k = 3$. Number of cointegrated relationships $r = 4$. The sample period is July 1, 2015, to December 4, 2018. Days with missing observations (i.e. non-trading days) were removed, resulting in a sample size of $n = 858$. Estimated, asymptotic standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[a]The volatility of bitcoin returns, σ^2, was scaled by multiplying each data point with 100 to avoid small coefficients.

C   VECM – Model Specifications

Step 1: Stationarity of variables. We first test the variables for unit roots and determine if the variables are stationary. We perform both the Augmented Dickey-Fuller (ADF) unit root test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) stationarity test. In the ADF test, the null hypothesis is that a variable contains a unit root, which indicates that it follows a non-stationary process. If the series is stationary after differencing once, it is integrated of order 1, i.e. $I(1)$. The alternative hypothesis is that the series is generated by a stationary process, and is thus integrated of order zero, i.e. $I(0)$. In the KPSS test, the null hypothesis is that the variable follows a stationary process, and is integrated of order zero. The alternative hypothesis is that the series is generated by a non-stationary process. If the series is stationary after differencing once, it is integrated of order one.

When performing the ADF test, we include a lag of five. This number represents the five trading days per week and ais thought to appropriately capture intra-week variations. As Table C1 shows, we reject the null hypothesis of a unit root for σ^2 and vix. These variables are thus stationary. We fail to reject the null hypotheses for the remaining variables and, after testing for higher order of integration, conclude

Table C1: Results of ADF Unit Root Tests

Variables	Test Stat[a]	p-value[b]	Order of Integration[c]
$ln(P)$	-1.244	0.658	$I(1)$
σ^2	-3.500	0.009	$I(0)$
V	-1.084	0.725	$I(1)$
IA	-1.668	0.449	$I(1)$
$sp500$	-0.917	0.783	$I(1)$
vix	-4.765	0.000	$I(0)$

Note. This table presents the results of the Augmented Dickey-Fuller (ADF) unit root tests. We used a vector autoregression (VAR) model with drift, but without time trend, and five lags representing the five trading days per week. The critical values are -3.453, -2.877, and -2.578 at the 1%, 5%, and 10% significance level, respectively. Variables are as defined in Table 1 and IA is the investor attention PCA index. Source: Authors' calculations and values obtained from MATLAB.
[a]The test statistics do not follow the usual t-distribution under the null hypothesis, since the null is one of non-stationarity, but rather follow a non-standard distribution (Brooks, 2014).
[b]Evaluated at the 5% significance level.
[c]$I(0)$ indicates that the time series is integrated of order zero. $I(1)$ indicates that the time series is integrated of order one.

Table C2: Results of KPSS Stationarity Tests

Variables	Test Stat[a]	p-value[b]	Order of Integration[c]
$ln(P)$	13.718	0.01	$I(1)$
σ^2	3.370	0.01	$I(1)$
V	13.430	0.01	$I(1)$
IA	9.287	0.01	$I(1)$
$sp500$	13.983	0.01	$I(1)$
vix	2.228	0.01	$I(1)$

Note. This table presents the results of the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) stationarity tests. We used a vector autoregression (VAR) model with drift, but without time trend, and five lags representing the five trading days per week. The critical values are 0.739, 0.463, and 0.347 at the 1%, 5%, and 10% significance level, respectively. Variables are as defined in Table 1 and IA is the investor attention PCA index. Source: Authors' calculations and values obtained from MATLAB.
[a]The test statistics are computed using an ordinary least squares (OLS) regression.
[b]Evaluated at the 5% significance level.
[c]$I(0)$ indicates that the time series is integrated of order zero. $I(1)$ indicates that the time series is integrated of order one.

that these time-series exhibit one unit root.

When performing the KPSS test, we also include a lag of five. As Table C2 shows, we fail to reject the null hypothesis for all variables. After testing for higher order of integration, we conclude that all time series are integrated of order one. Due to conflicting results, we proceed to test for autocorrelation. After visually displaying the time series' autocorrelation function, it is evident that all variables exhibit autocorrelation up to lag 20, see Figure C1. After taking the first difference once, none of the time series appear to exhibit autocorrelation, see Figure C2. Thus, we favor the KPSS test results that suggest all time series to be integrated of order one.

Step 2: Number of lags. We use the Akaike Information Criterion (AIC) to choose the optimal lag length in the model. We estimate VAR models with lag length varying from 0 to 12 and compute the AIC. The AIC for a VAR model is defined as $-2L + 2(p + 2kp)$, where L is the log-likelihood, p is the number of coefficients, and k is the lag length. A smaller AIC indicates better trade-off between model fit and complexity. Based on the results in Table C3, we select a

Table C3: Selecting Optimal Lag Length Using KPSS Test Results

Lag	AIC	BIC	HQIC
0	-37.276	-37.276	-37.276
1	-37.616	-37.418	-37.540
2	-37.700	-37.302	-37.548
3	-37.775^*	-37.179	-37.547
4	-37.765	-36.970	-37.460
5	-37.732	-36.739	-37.352
6	-37.717	-36.524	-37.260
7	-37.706	-36.315	-37.173
8	-37.702	-36.111	-37.093
9	-37.683	-35.895	-36.999
10	-37.658	-35.670	-36.897
11	-37.615	-35.429	-36.778
12	-37.577	-35.192	-36.664

Note. This table presents the results of using AIC to choose the optimal lag length in the model. AIC denotes Akaike Information Criterion, BIC is the Bayesian Information Criterion, and HQIC the Hannan-Quinn Information Criterion. We estimate vector autoregression (VAR) models with lag length varying from 0 to 12, and choose the model with the optimal trade-off between model fit and complexity. A lower AIC indicates a better model fit. Here, the optimal lag length is three, indicated by *. Source: Authors' calculations and values obtained from MATLAB.

VAR model with $k = 3$.[C1]

Step 3: Cointegration tests. We use the Johansen trace test to test for cointegration rank. The trace test is a sequential, joint hypothesis testing procedure where the null hypothesis is that the number of cointegrating vectors is less than or equal to rank (r), against an unspecified or general alternative that there are more cointegrating vectors than r. The test is repeated until the first null hypothesis is not rejected. From Table C4, we reject the first null hypothesis of no cointegration, which confirms that VECM is the appropriate model. The trace test stops at the null hypothesis when there are three cointegration relations in the bitcoin market. Therefore, we proceed to estimate the VECM with $r = 3$.

[C1]We also used the ADF test results to estimate VAR models. When doing so, the optimal lag length proved to be $k = 4$. Thus, the KPSS and ADF test provide conflicting results. However, due to the presence of autocorrelation, and to reduce dimensionality, we favor the KPSS results ($k = 3$).

Table C4: Johansen Trace Test for Cointegration

Rank	Eigenvalue	Trace Stat	Critical Value	p-value[a]
0	0.092	270.584	169.602	0.001
1	0.074	187.937	134.681	0.001
2	0.055	122.307	103.848	0.002
3	0.032	74.121*	76.972	0.081

Note. This table presents the results from the Johansen trace test for cointegration rank. The applied test assumes intercepts in the cointegrated series, no deterministic trends in the levels of the data, and lag length three (based on the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test results). The Johansen trace test stops at the null hypothesis when there are three cointegrating vectors, indicated by *. Source: Authors' calculations and values obtained from MATLAB.

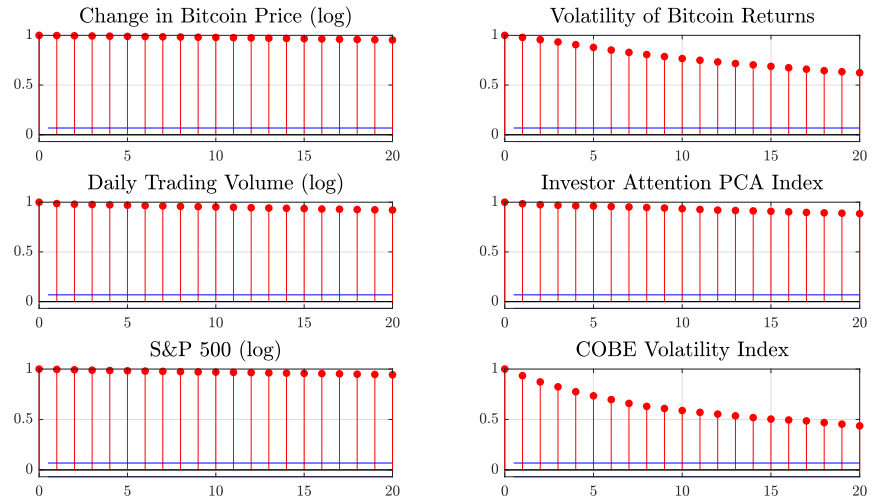[a]Evaluated at the 5% significance level.

Figure C1: Autocorrelation Function on Original Data Set

Note. This figure depicts the autocorrelation function for all the variables included in the VECM. Autocorrelation represents the degree of similarity between a given time series and a lagged version of itself over successive time intervals, i.e. the relationship between a variable's current and past values (Brooks, 2014). As can be seen, all variables exhibit autocorrelation up to lag 20.
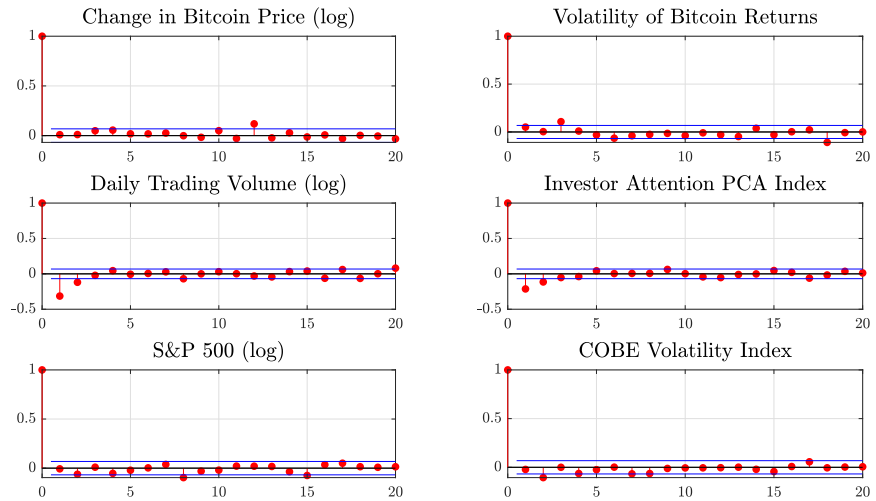
Figure C2: Autocorrelation Function on Differenced Data Set

Note. This figure depicts the autocorrelation function for all the VECM variables after taking first difference of all the variables. Autocorrelation represents the degree of similarity between a given time series and a lagged version of itself over successive time intervals, i.e. the relationship between a variable's current and past values (Brooks, 2014). As can be seen, after taking first difference once, none of the variables exhibit autocorrelation.

D VECM – Robustness Tests

Table D1: VECM Error Correction Estimates with Only Google Trends SVI

Independent Variables	Dependent Variables		
	$\Delta ln(P)_t$	$\Delta \sigma_t^2$	ΔV_t
$\Delta ln(P)_{t-1}$	-0.003	0.21	0.013^*
	(0.036)	(0.064)	(0.007)
$\Delta \sigma_{t-1}^2{}^{\text{a}}$	-0.041^{**}	-0.032	0.020^{***}
	(0.019)	(0.034)	(0.004)
ΔV_{t-1}	0.611^{***}	-0.880^{**}	-0.378^{***}
	(0.218)	(0.396)	(0.043)
$\Delta googletrends_{t-1}$	0.159	-0.880^{***}	0.047
	(0.147)	(0.267)	(0.029)

Note. This table presents the first lag coefficients of the short-term relationship between the model variables including only the Google Trends SVI as an investor attention metric. Estimates for controls are not displayed. Variables are as defined in Table 1. Lag length $k = 3$. Number of cointegrated relationships $r = 3$. The sample period is July 1, 2015, to December 4, 2018. Days with missing observations (i.e. non-trading days) were removed, resulting in a sample size of $n = 858$. Estimated, asymptotic standard errors are in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[a]The volatility of bitcoin returns, σ^2, was scaled by multiplying each data point with 100 to avoid small coefficients.

Table D2: VECM Error Correction Estimates with Only Wikipedia

Independent Variables	Dependent Variables		
	$\Delta ln(P)_t$	$\Delta \sigma_t^2$	ΔV_t
$\Delta ln(P)_{t-1}$	0.001	-0.001	1.114
	(0.001)	(0.001)	(6.365)
$\Delta \sigma_{t-1}^2$ [a]	0.001	0.001**	-27.014
	(0.003)	(0.001)	(33.400)
ΔV_{t-1}	0.004	0.006***	-84.867**
	(0.003)	(0.002)	(39.217)
$\Delta wikipedia_{t-1}$	-0.007***	0.004**	-28.297
	(0.002)	(0.002)	(29.153)

Note. This table presents the first lag coefficients of the short-term relationship between the model variables including only the daily Wikipedia search queries for bitcoin as an investor attention metric. Estimates for controls are not displayed. Variables are as defined in Table 1. Lag length $k = 3$. Number of cointegrated relationships $r = 3$. The sample period is July 1, 2015, to December 4, 2018. Days with missing observations (i.e. non-trading days) were removed, resulting in a sample size of $n = 858$. Estimated, asymptotic standard errors are in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
[a]The volatility of bitcoin returns, σ^2, was scaled by multiplying each data point with 100 to avoid small coefficients.

Table D3: VECM Error Correction Estimates with Only New Members

Independent Variables	Dependent Variables		
	$\Delta ln(P)_t$	$\Delta \sigma_t^2$	ΔV_t
$\Delta ln(P)_{t-1}$	−0.009	0.022	0.016**
	(0.035)	(0.064)	(0.007)
$\Delta \sigma_{t-1}^2$ [a]	−0.047**	−0.018	0.023***
	(0.019)	(0.034)	(0.004)
ΔV_{t-1}	0.623***	−0.754	−0.348***
	(0.216)	(0.392)	(0.042)
$\Delta newmembers_{t-1}$	0.367**	0.133	0.027
	(0.181)	(0.329)	(0.035)

Note. This table presents the first lag coefficients of the short-term relationship between the model variables including only the daily new members on Bitcointalk.org as an investor attention metric. Estimates for controls are not displayed. Variables are as defined in Table 1. Number of cointegrated relationships $r = 3$. The sample period is July 1, 2015, to December 4, 2018. Days with missing observations (i.e. non-trading days) were removed, resulting in a sample size of $n = 858$. Estimated, asymptotic standard errors are in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[a] The volatility of bitcoin returns, σ^2, was scaled by multiplying each data point with 100 to avoid small coefficients.

Table D4: VECM Error Correction Estimates with PCA Index and Dummy Variable

Independent Variables	Dependent Variables			
	$\Delta ln(P)_t$	$\Delta \sigma_t^2$	ΔV_t	ΔIA_t
$\Delta ln(P)_{t-1}$	-0.009	0.036	0.017^{**}	-0.012^{*}
	(0.001)	(0.063)	(0.007)	(0.007)
$\Delta \sigma_{t-1}^2$ [a]	-0.041^{**}	-0.024	0.024^{***}	-0.006
	(0.018)	(0.033)	(0.004)	(0.004)
ΔV_{t-1}	0.689^{***}	-0.933^{**}	-0.343^{***}	-0.086^{**}
	(0.216)	(0.388)	(0.042)	(0.042)
ΔIA_{t-1}	0.581^{***}	-0.489	0.093^{**}	-0.296^{***}
	(0.193)	(0.347)	(0.038)	(0.038)
$\Delta sp500_{t-1}$	0.010	-0.015	0.001	-0.001
	(0.006)	(0.012)	(0.001)	(0.001)
Δvix_{t-1}	-1.360	1.392	0.007	0.077
	(1.248)	(2.243)	(0.245)	(0.245)

Note. This table presents the estimated error correction terms and first lag coefficients of the short-term relationship between the model variables. $ln(P)$ is the bitcoin price (log), σ^2 is the volatility of bitcoin returns, V is the daily trading volume (log), and IA is the investor attention PCA index. Other variables are as defined in Table 1. For concreteness, the control variables are not displayed among the dependent variables. Lag length $k = 3$. Number of cointegrated relationships $r = 3$. The sample period is July 1, 2015, to December 4, 2018. Days with missing observations (i.e. non-trading days) were removed, resulting in a sample size of $n = 859$. Estimated, asymptotic standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
[a]The volatility of bitcoin returns, σ^2, was scaled by multiplying each data point with 100 to avoid small coefficients.