What Drives Bitcoin Volatility?

Byström, Hans; Krygier, Dominika

2018

Document Version:
Other version

Link to publication

Citation for published version (APA):
Byström, H., & Krygier, D. (2018). *What Drives Bitcoin Volatility?* (Working Papers; No. 2018:24).

Total number of authors:
2

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Working Paper 2018:24

Department of Economics
School of Economics and Management

What Drives Bitcoin Volatility?

Hans Byström
Dominika Krygier

October 2018

What Drives Bitcoin Volatility?

HANS BYSTRÖM and DOMINIKA KRYGIER*

We look at the link between the volatility in the Bitcoin market and the volatility in other related traditional markets, i.e. the gold, currency and stock market. We also try to answer if the volatility in the Bitcoin market can be explained by retail investor-driven internet search volumes or, perhaps, by the general level of risk in the financial system, as measured by two market-wide risk indicators. We use daily, weekly as well as monthly data covering the period 2011 to 2017. Correlations and regressions reveal a weak but positive contemporaneous link between changes in the Bitcoin volatility and changes in the volatility of the trade weighted USD currency index. A stronger positive link is found between Bitcoin volatility and search pressures on Bitcoin-related words on Google, particularly for the word "bitcoin". To further assess what drives Bitcoin volatility we turn to a VAR-analysis and impulse response functions which point at Google searches for the word "bitcoin", and to some extent the USD currency index volatility, being the only determinants of future Bitcoin volatility. We then use our findings to make improved predictions of Bitcoin volatility based on Google search activity. Interestingly, the significant link that we find between Google search volumes and market volatility points at retail investors, rather than large institutions, being the most important drivers of Bitcoin volatility. We believe that we contribute to the literature in several ways and that our results could be of significant practical importance if the Bitcoin market continues to grow at the current speed.

Keywords: Bitcoin; volatility; internet searches; Google Trends; gold; VIX
JEL classification codes: G10; D80; C80

* Hans Byström is from the Department of Economics, Lund University, Box 7082, 22007 Lund, Sweden (hans.bystrom@nek.lu.se) and Dominika Krygier is from the Department of Economics, Lund University, Box 7082, 22007 Lund, Sweden (dominika.krygier@nek.lu.se). The authors thank Google for making the search volume publicly available through Google Trends. All mistakes in this article are the authors'.

Introduction

One of the most intriguing financial innovations of the last decade is without doubt the concept of cryptocurrencies. Among the many existing cryptocurrencies the most well-known one is Bitcoin. The market capitalization of the entire cryptocurrency market, as of January 2018, is around $bn700 and the Bitcoin market makes up roughly one third of that market (Coinmarketcap.com, 2018). Although the Bitcoin market is dwarfed by many traditional financial markets, such as the stock market which has a market capitalization of close to $tn100 or the gold market with a market capitalization somewhere in the $tn10 - $tn100 range, the Bitcoin market is currently growing quickly. One indicator of the scale of the Bitcoin market is the electricity consumption needed to keep the cryptocurrency market alive. Estimates of the electricity needs vary widely from the energy production of a large nuclear reactor to the energy consumption of a small industrialized country such as Denmark (Bloomberg, 2017).

In this paper we try to answer whether the volatility in the Bitcoin market can be explained by volatility levels in traditional financial markets dominated by institutional investors or, perhaps, by internet search activity, which is thought to be created mainly by retail investors and the general public. We also ask whether the Bitcoin volatility perhaps can be explained by the general level of risk in the financial system, as measured by two different risk indicators. In addition to the general academic interest in explaining what drives the price movements in a novel financial market such as the Bitcoin market, there are several practical reasons for looking into this issue. In fact, a range of new innovations in the Bitcoin market highlights the need for more knowledge about causes and features of Bitcoin volatility. One example is the launch in 2015 of a Bitcoin version of the VIX fear index, the so-called .BVOL index, by the Bitcoin derivatives exchange BitMEX (Wong, 2014). The .BVOL index is an index of the volatility in the Bitcoin market, comparable in spirit to the VIX stock volatility index, and it is available in daily, weekly and monthly versions. Another example is the recent introduction of Bitcoin futures trading on the two derivatives exchanges Chicago Mercantile Exchange (CME) and Chicago Board Options Exchange (CBOE). To compute the margins required by clearing houses and brokers standing behind the buyers and sellers in such Bitcoin futures markets one need to make predictions of the Bitcoin volatility (Financial Times, 2017a). Furthermore, the launch of Bitcoin futures by two of the world's major derivatives exchanges has also led to several firms trying to get approval for Bitcoin-tracking exchange traded funds (ETFs) that track futures prices instead of spot prices (Financial Times, 2017b). Currently, there are very few asset classes for which there are no ETFs and an introduction of Bitcoin ETFs could further spur the development of the Bitcoin market. Such a development, with a widening range of potential Bitcoin investors, creates a growing need for a deeper understanding of risk and volatility in the Bitcoin market.

We look at Bitcoin prices and how the volatility of Bitcoin returns is linked to corresponding volatilities in the gold, currency and stock market as well as to the level of risk in the financial system, measured by two market wide risk measures. We also link the Bitcoin volatility to Google internet search volumes on phrases like "bitcoin", "gold price", "war" and "cyber attack" using Google Trends. The linkages are studied using data sampled on a daily, weekly as well as monthly frequency, and the time period begins in 2011, when a liquid secondary market for Bitcoins had developed, and ends in 2017.

Correlations and OLS-regressions reveal a positive link between contemporaneous changes in Bitcoin volatility and the USD trade weighed currency index volatility (USD volatility from now on) but the most significant link is found between Bitcoin volatility and search pressures on Bitcoin-related words on Google. To study lead-lag relationships between the variables, and to assess the ability to predict Bitcoin volatilities, we also turn to formal VAR-analysis and impulse response functions, and the results point at Google searches for the word

Bitcoin, and to some extent USD volatility, being the only statistically significant determinants of future Bitcoin volatility. Finally, we use our findings to predict Bitcoin volatility out-of-sample. When evaluated using various volatility forecasting evaluation methods we find that, overall, when predicting volatilities in the Bitcoin market it is worthwhile acknowledging search pressure on search engines like Google.

We believe that we contribute to the literature in several ways. First, it is (one of) the first academic studies to look into the causes of the (very high) volatility in the young but quickly growing Bitcoin market.[1] Second, we look at volatilities at different sampling frequencies, with both daily, weekly and monthly windows for volatility calculations. Third, by looking at Google search volumes we believe that we can isolate, at least to some degree, the share of the driving forces behind Bitcoin volatility that are related to the retail market. Bitcoin is often regarded as (merely) a speculative tool for retail investors (Financial times, 2017c; MotleyFool, 2017) and if it is true that the market behavior is shaped by these retail investors then there are reasons to believe that the volatility in the market would be primarily caused by retail investors. In fact, the significant positive link that we find between Google search volumes and Bitcoin volatility supports this; i.e. that retail investors, rather than large institutional investors, are the fundamental drivers of Bitcoin volatility.

The rest of the paper is organized as follows. In Section I we give a brief description of the Bitcoin market and in Section II we review the literature. Section III describes the data. In Section IV we present empirical evidence on the drivers of Bitcoin volatility. Section V, finally, concludes.

---

[1] The only other similar study of the bitcoin market that we are aware of is Urquhart (2018), who focuses on "what factors drive the attention of Bitcoin?". Like us, Urquhart (2018) employs Google Trends data as a proxy for investor attention but then focuses on whether realized volatility (as well as returns or volume) are significant drivers of the attention of Bitcoin. As a side result, however, Urquhart (2018) finds that investor attention offers no significant predictive power in forecasting realized volatility; "the estimation results also reveal that past search queries does not significantly influence realized volatility as the coefficient is only significant at the 10% level".

I. Bitcoins and the Bitcoin Market

Bitcoin is a cryptocurrency, or more exactly a digital cash peer-to-peer network, that works without a central authority for settlement and validation of currency transactions. Like all cryptocurrencies, Bitcoins have no underlying assets, are not backed by any government and pay neither interest nor dividends. There is no government (i.e. no central bank) backing the currency and Bitcoins are instead issued through a process called mining where miners provide necessary processing power to the Bitcoin network in exchange for Bitcoins. Bitcoin was introduced in October 2008 by Satoshi Nakamoto, which is thought to be an alias, and the key innovation is the way the decentralized Bitcoin network solves the so-called double-spend problem that digital currencies typically suffer from (preventing a certain Bitcoin to be spent more than once by the current owner). Bitcoin transactions are validated by a network of nodes (the miners) that verify the accuracy of every transaction using previous transactions registered in a ledger called the blockchain. Any new transaction is subsequently added to the blockchain and verified by the entire decentralized network through a concept called proof-of-work (Antonopoulos, 2017; Baur & Dimpfl, 2017). As a result of the mining process, the number of Bitcoins in circulation is steadily increasing. The total number of Bitcoins, however, is capped by the Bitcoin computer algorithm at 21 million, making the Bitcoin deflationary, rather than inflationary like traditional currencies. As of January 2018, close to 17 million Bitcoins are in circulation (i.e. 80 % of the hard-limit total money supply which will be reached in year 2140) and the total market value is around $230 billion (Blockchain.com, 2018, Coinmarketcap.com, 2018).

The identity of the typical Bitcoin user is not fully known. The encryption technology behind Bitcoin promises the user (more or less) anonymity. According to Yelowitz & Wilson (2015), however, anecdotal evidence puts the Bitcoin user into one of four clienteles; “computer programmer enthusiasts”, “speculative investors”, “libertarians” and “criminals”. Computer wizards are attracted by the possibility to earn money through faster and better mining than their competitors, speculators are tempted by the high volatility of the Bitcoin price, libertarians like the idea of bypassing central authorities such as central banks, and criminals appreciate the near-anonymity of Bitcoin transactions. Yelowitz & Wilson (2015) uses Google Trends to analyze the clientele effect in the Bitcoin market and among the four groups identified above, Yelowitz & Wilson (2015) only finds computer enthusiasts and criminals to be behind the (search query) interest in Bitcoin.

A related question is what/who determines the Bitcoin price. Is the price driven by internal forces, i.e. by the Bitcoin market participants themselves, or by external forces such as macro-variables or prices in other financial markets? Baek & Elbeck (2015) finds no significant link between the latter (external fundamental economic factors) and changes in Bitcoin prices, and conclude that Bitcoin returns are driven mainly by the Bitcoin buyers and sellers themselves. They interpret this as evidence of the Bitcoin market being an early-stage market with highly speculative features.

II. Literature Review

Few studies have been conducted on the characteristics of Bitcoin volatility; focus has instead been on the price formation and on the main drivers of the Bitcoin price. Three groups of explanatory variables have typically been used; variables related to investor sentiment and attention, variables related to Bitcoin supply and demand, and variables related to macro-finance. Surprisingly rarely, variables from the various groups have been combined in a single study. Generally, however, empirical research shows that the price formation is due to factors that substantially differ from those affecting conventional assets. Some of these factors include internet search and social media activity (Kristoufek (2013, 2014), Garcia et al. (2014), Kaminski

(2014)) as well as Bitcoin trade volume and supply (Balcilar et al. (2017)), with the former being the most studied one. As for volatility, the explanatory power of these factors has been varying, with perhaps internet search and social media activity being the most consistent ones when it comes to both predicting and explaining Bitcoin volatility.

Further, the status of Bitcoin as an asset has been hard to define – is it a currency or is it a commodity? It has been compared to gold and the US dollar, as it shares many similarities to both assets. Both gold and Bitcoin are costly to obtain, neither of them is controlled by a nation or a government and both assets are extracted through the process of "mining" by independent operators. Gold has long been considered a safe haven asset and a hedging instrument because of its negative correlation with the US dollar. Bitcoin, however, has not typically been recognized as a hedge (Dyhrberg (2016), Bouri et al. (2017)). Similar arguments can be used for the similarity between Bitcoin and the US dollar. Both have no or limited intrinsic value, but while the US dollar is backed by the government Bitcoin is not. Accordingly, some, like Yermack (2015), argue that Bitcoin largely fails to satisfy the criteria for being a fiat currency.

Ciaian et al. (2016) studies the Bitcoin price formation by considering both traditional determinants of currency prices and Bitcoin, or digital currency specific factors. They find that market forces of supply and demand, mainly Bitcoin trading volume, number of Bitcoins outstanding and price level, have strong impact on price formation. Further, they find no evidence of the conjecture that macro-financial developments, such as for example stock exchange indices, commodity prices or inflation, should drive the Bitcoin price in either the short or the long run. Kristoufek (2013) investigates the relationship between the Bitcoin price and investors' interest and attention using search queries on Google and Wikipedia as proxies. He finds a strong correlation between the price and search queries on both internet platforms. However, when prices are above trend, the increasing interest in Bitcoin leads to a continuation of the rise in the price, and the other way around when prices are below trend. This bi-directional, and asymmetric, relationship is argued to be a common sight for financial assets with no underlying fundamentals, such as Bitcoin. In a later paper, Kristoufek (2015) extends the analysis by studying possible fundamental, or economic, drivers, followed by transactional drivers (the use of Bitcoin in real transactions) and technical drivers (the mining process). He finds that Bitcoin behaves according to standard economic theory in the long run, but is prone to bubbles and busts in the short run. From a technical standpoint, when the Bitcoin price increases, users are motivated to start mining. Kristoufek (2015) also finds no signs of Bitcoin being a safe-haven asset, a hypothesis that has been explored also by others.

When it comes to Bitcoin volatility, the list of potential drivers that have been tested is more or less the same as the drivers of the Bitcoin price, at least when observing the research output. Bouri el al. (2016) model the Bitcoin volatility by applying the asymmetric GARCH model in order to test the impact of positive and negative shocks (news). They do find a positive relationship between shocks to return and volatility, but only in the pre-crash period (up until June 20, 2011 when Mt. Gox was exposed to hackers resulting in a price dip of Bitcoin to 0.01 US dollar in only a couple of minutes). Bouri et al. (2016) also find a negative relationship between the VIX index and the Bitcoin realized volatility. Dyhrberg (2016) applies the GARCH framework to analyze the behavior of Bitcoin volatility in comparison to gold and the US dollar-euro exchange rate. Similar to gold, the price volatility of Bitcoin also exhibits volatility clustering and high volatility persistence. Dyhrberg (2016) also finds that past volatility as a predictor for future volatility dominates the predictive ability of shock (news) effects for Bitcoin.

A study similar to ours is that of Urquhart (2018) who focuses on “what factors drive the attention of Bitcoin?”. Like us, Urquhart (2018) employs Google Trends data as a proxy for investor attention but then focuses on whether realized volatility (as well as returns or volume) are significant drivers of the attention of Bitcoin. As a side result, however, Urquhart (2018) finds that investor attention offers no significant predictive power in forecasting realized volatility. Instead, the previous day volatility and volume, as well as two days previous returns are found as significant drivers of the attention of Bitcoin. This result holds, as the authors states, from October 2013. Urquhart suggests that this might be due to the fact that investors are attracted to Bitcoin after increases in volatility and trading volume. Urquhart (2018) studies a time period that ends just one month after ours but compared to us they only look at daily data and US-based Google searches.

Overall, there is a strong indication that the Bitcoin price and volatility dynamics are influenced by social factors connected to internet search activities. Other currencies and commodities, such as gold and oil, also seem to play a role, perhaps motivated by the finding that Bitcoin is somewhere between being a currency and a commodity. Our paper contributes to the existing literature by combining social factors, general macro-financial risk measures, and volatilities in other traditional financial markets to investigate the driving forces behind the Bitcoin volatility. We believe that an interaction of the said three groups of possible driving forces may be informative when studying the volatility of Bitcoin.

III. Data

The data used in this paper covers the time period August 2011 to June 2017 and is sampled on a daily, weekly and monthly frequency.[2] The Bitcoin price data (USD/Bitcoin) is downloaded from Datastream and is originally sourced from the Luxembourg-based Bitcoin exchange Bitstamp. Gold price data (USD/Oz), stock price index data (S&P500 Composite Index) as well as USD currency index data (a trade-weighted USD index) is also downloaded from Datastream. The two

[2] The data for the two higher frequencies covers the longest possible single continuous sub-periods that we manage to construct using the Google Trends downloading mechanism (on July 21, 2017). The prices are all end-of-day quotes.

risk measures, a global economic policy uncertainty index (EPU) and a systemic risk indicator, are from Baker, Bloom and Davis (2016), and the Federal Reserve Bank of Cleveland, respectively. Like the price data described above the risk measure data covers the time period August 2011 to June 2017 and is sampled on a daily, weekly and monthly frequency.

The EPU index is based on three components; nation-wide newspaper coverage frequency of words connected to: the economy (E), policy (P) and uncertainty (U), temporary federal tax code provisions that are set to expire within the next 10 years, and disagreement among economic forecasters, where the dispersion between individual forecasters' prediction about future macro-economic variables (such as CPI, federal expenditures, state expenditures) is used to construct an uncertainty index about overall policy-related macro-economic variables. These three components are used to capture overall policy-related economic uncertainty within a country, or globally. For a detailed description of the index see Baker, Bloom and Davis (2016).

The systemic risk indicator measures systemic risk in the US banking industry and is provided by the Federal Reserve Bank of Cleveland (originating from Saldías (2013)). The index is based on calculating an insolvency measure that is centered both on individual banking institutions and financial intermediaries, as well as the banking system seen as a whole. The indicator is constructed to gauge market wide perceptions of the risk of widespread insolvency in the banking system. Details on the indicator can be found on the website of the Federal Reserve Bank of Cleveland, and in Saldías (2013).

The Google Trends data covers different time periods depending on the sampling frequency (see footnote 2); for the daily sampling frequency the time period is December 21, 2016 – June 19, 2017, for the weekly sampling the time period is June 24, 2012 – June 18, 2017, and for the monthly sampling the time period is August 2011 – June 2017. The Google Trends search volumes are downloaded (on July 21, 2017) for the nine search strings "bitcoin", "VIX", "crisis", "cyber attack", "gold price", "interest rate rise", "inflation", "stock market crash" and "war".[3] Google Trends started publicly releasing data on search term intensity in 2009. Rather than providing a measure that portrays the absolute number of searches, the search term intensity is set relative to all other searches over a certain period of time. By indexing the search interest in this way one controls for any change in the overall internet activity over time. In addition to this indexing the search intensity is also normalized to vary between 0 and 100 where the highest search intensity across the particular time period is set to 100.

Volatilities for the price changes (log-returns) of Bitcoins, US stocks, the USD dollar index and gold are calculated as ordinary weekly or monthly sample standard deviations for the weekly and monthly sampling frequencies (using daily price changes and subtracting the mean) and as squared price changes for the daily frequency. All Google Trends data is normalized as described above and expressed in relative terms.

IV. Econometric Method and Results

A. Descriptive Statistics and Correlation Analysis

Table 1-3 report descriptive statistics for our variables (log first differences). The first Table is for the daily frequency, where we have 113-128 observations depending on the variable (there are some missing values for some variables), the second Table presents the weekly statistics, where the number of observations is 260, and, finally, the third Table is for the monthly frequency, with

[3] Throughout the paper we use the following abbreviations for the various variables: the Bitcoin volatility (BTCVOL), the gold volatility (GOLDVOL), the USD index volatility (USDVOL), the US stock volatility (S&P500VOL), the economic policy uncertainty index (EPU) , the systemic risk indicator (SYS), Google volume for "bitcoin" (GBTC), Google volume for "VIX" (GVIX), Google volume for "crisis" (GCRISIS), Google volume for "cyber attack" (GCYBER), Google volume for "gold price" (GGOLD), Google volume for "interest rate rise" (GINTE), Google volume for "inflation" (GINFL) , Google volume for "stock market crash" (GCRASH) and Google volume for "war" (GWAR).

70 observations. All variables demonstrate a little bit of skewness as well as kurtosis. However, only a few of the Google search terms (different search terms for different frequencies) have high levels of kurtosis. Other than that the deviation from normality is not significant. As for the main variable, the Bitcoin volatility, Tables 1-3 show that the 'volatility of the volatility' is very high across all frequencies. So it is not only the volatility of the Bitcoin price that is high, which is a well-known stylized fact reported in the media (Financial Times, 2018), but the volatility itself fluctuates a lot as well.

We continue by investigating pairwise correlations between all our variables, for the three frequencies. Correlation matrices are found in Tables 4-6. Our focus is primarily on the correlations between the Bitcoin price volatility (change) and each of the other 14 variables, i.e. the first column in the Tables. To start with, we notice that the correlation coefficient rarely reaches levels above +/-0.5. Many correlation coefficients are also close to zero, and non-significant, indicating a non-existing or weak relationship. As for the Bitcoin volatility, the majority of the correlations with the other variables are positive, regardless of frequency. Half of the correlation coefficients are statistically significant, as indicated by the stars in the Tables, and most of the significant correlations are positive.

Among the Google variables, the Google search volume for the very word "bitcoin" appears to have some importance; the correlation coefficient is statistically significant for all frequencies. At this point we cannot draw any common conclusions regarding the general Google search terms and its correlation with Bitcoin volatility since the correlations are sometimes negative and sometimes positive depending on the frequency of the data. The significance level also varies among the various search terms and frequencies. Overall, however, most of the Google correlations are positive, particularly the statistically significant ones, so there seems to be a positive link between Google search pressure and Bitcoin price volatility.

As for how the Bitcoin volatility is correlated with volatilities in other markets, we observe that for all frequencies, the volatility of the gold price is positively correlated with the volatility of the Bitcoin price. The correlation coefficient decreases with a declining data frequency (15%*, 12%*, 7%) and is statistically significant for daily- and weekly data. The Bitcoin volatility also appears to be positively correlated with the volatility in the US currency market, while the link with the US stock market is inconclusive or non-existing. Overall, the gold price volatility shows the strongest correlation with Bitcoin volatility for our sample period.

The two risk measures (EPU and SYS) show mixed and statistically non-significant results, and we cannot reject the null hypothesis that the correlation coefficient between our two (well-known and widely used) risk measures and Bitcoin volatility is zero.

B. OLS Regression Analysis

The main question that we are trying to answer in this paper is what drives Bitcoin volatility. The simplest way to do this is by regressing Bitcoin volatility on a set of variables that we believe have the potential to explain and, more importantly in a later stage, forecast Bitcoin volatility. Our empirical model is hence a straightforward time-series OLS regression of possible determinants of the volatility in the Bitcoin market.

We begin by conducting unit-root tests on all our variables to make sure that they are all stationary. We start by studying the original data sample, i.e. in levels. For all frequencies, the null hypothesis of an augmented Dickey-Fuller (ADF) unit root test cannot be rejected. This is also the case for the logarithmic transformation of the data. Only when transforming the data by taking the logarithm of the first differences, do ADF tests, as well as simple plots of the data series, show that our variables are stationary. Consequently, all forthcoming analysis is based on logged first-differences of the time series data.

We perform a total of six ordinary least squares regressions, two for each data frequency. The first regression for each frequency contains all the variables (at the same point in time), while the second regression also includes (one period) lagged Bitcoin volatility (log difference) on the right hand side since yesterday's value is likely to help predict today's value and as a control for possible autocorrelation in the (logged first difference) Bitcoin volatility. To sum up, we estimate the following two models for each of the three frequencies:

$$BTCVOL_t = \beta_0 + \beta_1 GOLDVOL_t + \beta_2 USDVOL_t + \beta_3 SPVOL_t + \beta_4 EPU_t + \beta_5 SYS_t + \beta_6 GBTC_t + \beta_7 GVIX_t + \beta_8 GCRISIS_t + \beta_9 GCYBER_t + \beta_{10} GGOLD_t + \beta_{11} GINTE_t + \beta_{12} GINFL_t + \beta_{13} GCRASH_t + \beta_{14} GWAR_t + \varepsilon_t \quad (1)$$

$$BTCVOL_t = \beta_0 + \beta_1 GOLDVOL_t + \beta_2 USDVOL_t + \beta_3 SPVOL_t + \beta_4 EPU_t + \beta_5 SYS_t + \beta_6 GBTC_t + \beta_7 GVIX_t + \beta_8 GCRISIS_t + \beta_9 GCYBER_t + \beta_{10} GGOLD_t + \beta_{11} GINTE_t + \beta_{12} GINFL_t + \beta_{13} GCRASH_t + \beta_{14} GWAR_t + \beta_{15} BTCVOL_{t-1} + \varepsilon_t \quad (2)$$

The Bitcoin volatility is the dependent variable and as explanatory variables we have the gold volatility, the USD currency index volatility, the S&P500 composite stock index volatility, the Economic Policy Uncertainty (EPU) index, the systemic risk (SYS) indicator from the Federal Reserve, and our nine Google search volume variables: 'bitcoin', 'VIX', 'crisis', 'cyber attack', 'gold price', 'interest rate rise', 'inflation', 'stock market crash' and 'war'.[4] The time period is t=1...T, where the unit of time (and the length of the time-period T) is either one day (T=111), one week (T=260) or one month (T=70) depending on the sampling frequency.

In Table 7 we present the results from the OLS regressions. There are three main columns representing each of the three frequencies, and for each frequency there are two specifications, one excluding and one including the lagged Bitcoin volatility. All through, the regression

[4] See footnote 3 for the variable abbreviations.

parameter for the lagged Bitcoin volatility is highly significant and indicates a negative relationship between subsequent Bitcoin volatility changes.

We notice that the Google search term for the word "bitcoin" is significant in all specifications except the 4^{th}. The sign of the OLS parameter is not always positive, though; just like for the correlations the OLS-relationship is negative (but not consistently statistically significant) for weekly data. However, similar to the correlations, with data sampled at daily and monthly frequencies, the OLS-relationship between Google search volumes for the word "bitcoin" and the Bitcoin volatility is positive and highly significant. Among the other Google search terms, the results again resemble those for the correlations. The relationship is sometimes negative and sometimes positive depending on the frequency of the data. The significance level also varies among the various search terms and frequencies.

As for the link between Bitcoin volatility and the volatility in other more traditional financial markets, Bitcoin volatility is still positively and statistically significantly correlated with the volatility in the US currency market, while the link with volatilities in both the gold and the US stock market is now inconclusive or non-existing.

The two risk measures (EPU and SYS) show a negative relationship with the Bitcoin volatility for all frequencies. The relationship is not statistically significant, however, and we therefore cannot reject the null hypothesis that there is no relationship between the risk measures and the volatility in the Bitcoin market.

The monthly R^2's are lowest for the weekly frequency. This is possibly linked to the odd results for the Google Bitcoin search term when weekly data is used. Overall, R^2's are higher when lagged Bitcoin volatility is added to the regressions on the right-hand side.

C. Vector Auto Regressions and Impulse Response Function Analysis

We find no evidence of cointegration among the variables, hence in order to get a dynamic view of how Bitcoin volatility is affected by our variables we estimate an unrestricted vector autoregressive (VAR) model. In a VAR model we model the linear interdependence between multiple time series, where the dependent variable, as well as all the other variables, is regressed on lagged values of themselves, in a system. By doing this, we can study both the contemporaneous effect, and any possible lag effects of the other variables that drive the Bitcoin volatility. We evaluate the estimated VAR models by means of generalized impulse response functions.

To start with, we employ a bivariate VAR framework; that is, we estimate a bivariate VAR(p) model (where p is the number of lags) that includes the Bitcoin volatility and each one of the remaining 14 variables, for each frequency. Due to the large number of coefficients to be estimated we chose not to estimate VAR models that have more than two variables at this point. The bivariate choice is also based on the fact that we have a fairly small number of observations relative to the number of coefficients to estimate. This modeling, however, still gives us important information, as a complement to the OLS estimation, in the form of tracing out the bivariate relationship dynamics over time. We evaluate the results by estimating generalized impulse response functions that illustrate how the Bitcoin volatility reacts if a given explanatory variable is hit by a shock with a size equal to one standard deviation (in-sample effect).

The number of lags (p) to be included in the VAR models is determined by running a battery of tests.[5] Lag length selection is critical since long lag lengths wastes degrees of freedom and short lag lengths may lead to a misspecification of the model. In total, 6 lags are picked for

[5] The tests we use are: likelihood ratio, final prediction error, Akaike, Schwarz and Hannan-Quinn. We use several methods to investigate optimal lag lengths since some of the test-statistics are, for example, sensitive to testing in small samples (such as the likelihood ratio).

the daily data, 1-5 lags (depending on the variable) for the weekly data, and 3 lags for the monthly data. All the VAR models are stable, and we have no serial autocorrelation.

We estimate bivariate VAR(p) models, where p is separately determined from the information criteria previously mentioned, and where x_t^i $(i = 1, \ldots, 14)$ below represents each of the remaining 14 variables. The number of variables in the VAR(p) model is hence always equal to two, the Bitcoin volatility ($BTCVOL$) and each one of the other 14 variables (x_t^i). The following bivariate system is estimated:

$$BTCVOL_t = a_{BTC}^0 + a_{11}^1 BTCVOL_{t-1} + \cdots + a_{1p}^1 BTCVOL_{t-p} + a_{11}^2 x_{t-1}^i + \cdots + a_{1p}^2 x_{t-p}^i + e_{BTCVOL,t}$$
$$x_t^i = a_{x^i}^0 + a_{21}^2 x_{t-1}^i + \cdots + a_{2p}^2 x_{t-p}^i + a_{21}^1 BTCVOL_{t-1} + \cdots + a_{2p}^1 BTCVOL_{t-p} + e_{x^i,t} \tag{3}$$

As can be seen above, in a VAR(p) model with two variables we get one separate equation for each variable. Each equation then contains p lagged values of itself and p lagged values of the other variables. The resulting 2•(1+2p) coefficients are estimated by OLS. After estimation, we confirm that our estimated models are stable by observing that all inverse roots lie inside the unit circle.

We evaluate the results of our estimated VAR models by means of generalized impulse response functions (first proposed in Koop, Pesaran and Potter (1996) and further developed in Pesaran & Shin (1998)), as we are interested in how the Bitcoin volatility reacts to shocks in one of our other variables. We deviate from the traditional Sims' (1980) orthogonalized impulse response functions due to their dependency on the ordering of the variables in the VAR system. We have no way of establishing a clear, or economically motivated, ordering of our variables. We then assess the response of the Bitcoin volatility (expressed in logged first differences) to a positive one standard deviation shock in each of our explanatory variables. The impulse response function shows us the dynamic response to this shock of the Bitcoin volatility variable, with a 95% confidence interval calculated using Monte Carlo methods with 1000 iterations. We can preliminarily asses the impulse response functions by observing the eigenvalues, or inverse roots, of the VAR systems. All of them are less than unity which means that our VAR models are stable and, accordingly, that the resulting impulse response functions should decay over time.

Figures 1a,b-3a,b present 14 impulse response functions (and corresponding accumulated responses) each of the Bitcoin volatility, when every one of the included variables is shocked (increased) by one standard deviation in the residuals, based on the bivariate VAR models. For many of the variables, the impulse response function indicates a statistically non-significant impact in the first period following the shocks. In the subsequent periods, the response then either decays or oscillates slowly over time. The only impulse response that is statistically significant for all three data frequencies in the first period is the one for the Google variable GBTC (Google search intensity for the word 'bitcoin'). Again, the response is a positive one for daily and monthly data but a negative one for weekly data.

Observing the impulse response function for the daily frequency in Figure 1a, we see that when GBTC is hit by a one standard deviation shock in the residuals, the Bitcoin volatility increases 0.4% the day following the shock. The accumulated response in Figure 1b, although largely positive across the entire post-period, is not statistically significant. The same applies for the weekly and monthly frequencies in Figures 2-3 with a peak impulse response of, respectively, -0.08% the first week after the shock and 0.3% the first month after the shock. Like for the daily frequency, the accumulated response is overall positive for the entire post-period, albeit not statistically significant, for both the weekly and the monthly frequency.

The only other financial market where a shock to the volatility seems to have an effect on the volatility in the Bitcoin market is the USD currency market. The impulse response is positive for all frequencies and significant for both weekly and monthly data. The result is strongest for monthly volatilities and when the USD volatility is hit by a one standard deviation shock in the residuals, the Bitcoin volatility increases by 0.23% the month following the shock. These findings strengthen the previous results where both correlations and regressions indicates a positive link between volatilities in the two (currency) markets for Bitcoin and the US dollar. The impulse response results point at increased USD volatility spilling over to increased Bitcoin volatility.

In addition to the bivariate VAR-models described above, where all 14 explanatory variables are included one-by-one, we also estimate multivariate VAR models for the subset of explanatory variables that have significant coefficients in the multivariate OLS regression in Section IV B. These results can be found in Figures 4-6.

For the daily frequency (figure 4) we estimate a VAR(2) model with five explanatory variables, for the weekly frequency (figure 5) we estimate a VAR(2) with seven explanatory variables and for the monthly frequency (figure 6) we estimate a VAR(1) with four explanatory variables. The impulse response results are similar to those for the bivariate VAR estimations above. The Google variable GBTC again has a more significant and consistent effect on the Bitcoin volatility than the other variables, and, again, the response is a positive and statistically significant one for daily and monthly data but a negative (and non-significant) one for weekly data. When the variable GBTC is hit by a one standard deviation shock in the residuals, the Bitcoin volatility increases 0.4% (0.35%), respectively, the day (month) following the shock. The effect for the weekly frequency is not significant (-0.06%). These results are essentially the same as those for the bivariate VAR models.

Another more or less unchanged result is that the only other financial market affecting the volatility in the Bitcoin market is the USD currency market. The impulse response is weaker than for the bivariate VAR but it is still positive for all frequencies (i.e. weekly and monthly) where it is included in the multivariate VAR. Like for the bivariate VAR, the result is strongest for monthly volatilities and when the USD volatility is hit by a one standard deviation shock in the residuals, the response of the Bitcoin volatility the first month is 0.25% while the accumulated response is 0.15%.

D. Out-of-Sample Forecasting

The results from the regressions in Section B can potentially be useful for forecasting purposes. The only variable that is significantly related to Bitcoin volatility, however, is the Google search volume for the word "bitcoin". Through the regressions (Eq. 1-2) we show that there is a contemporaneous positive *relationship* between the two variables. The next step is to investigate whether Google can be used to *predict* Bitcoin volatility as well. Being able to forecast the volatility in such a volatile market as the Bitcoin market is, of course, important both in risk management and in trading situations.

As benchmark predictive models for volatility changes we use two naïve predictors; the *random walk-model* (the predicted next-period volatility is assumed to be unchanged from the current-period volatility) and an *AR-model* (the predicted next-period volatility is assumed to be related to the current-period volatility through the auto-regressive relationship in Eq. 2). In order to make the most out of our model, which we call the *Google-enhanced model,* we construct those predictions using both the current-period GBTC variable and current-period volatility since these two variables are the only ones that are significant at the 1% level in the OLS-regression

(Table 7).[6] Regardless of model, the predicted future volatility change is transformed to a predicted future volatility (i.e. level) labeled $\sigma_{forecast,t}$ which is then compared to the actual future volatility (i.e. variance). For the forecasting evaluation the time period is divided into an estimation period (expanding window) and an out-of-sample period. The out-of-sample period is one month for the daily frequency and one year for the weekly and monthly frequencies.

The forecasting performance is assessed using several different loss functions; the root mean squared error (*RMSE*), the percentage squared error (*PSE*), the quasi-likelihood (*QL*) loss function and the *R^2LOG* loss function. These are the loss functions suggested by Brownlees, Engle and Kelly (2011) and Bollerslev, Engle and Nelson (1994) and they are defined as

$$RMSE = \sqrt{\frac{1}{n}\sum_{1}^{n}\left(\sigma^2_{realized,t+1} - \sigma^2_{forecast,t}\right)^2} \tag{4}$$

$$PSE = \frac{1}{n}\sum_{1}^{n}\frac{\left(\sigma^2_{realized,t+1} - \sigma^2_{forecast,t}\right)^2}{\sigma^4_{forecast,t}} \tag{5}$$

$$QL = \frac{1}{n}\sum_{1}^{n}\left(\frac{\sigma^2_{realized,t+1}}{\sigma^2_{forecast,t}} - \log\left(\frac{\sigma^2_{realized,t+1}}{\sigma^2_{forecast,t}}\right) - 1\right) \tag{6}$$

$$R^2LOG = \frac{1}{n}\sum_{1}^{n}\left(\log\left(\frac{\sigma^2_{realized,t+1}}{\sigma^2_{forecast,t}}\right)\right)^2 \tag{7}$$

where $\sigma_{forecast,t}$ is the volatility forecast produced at day/week/month *t* of the volatility in day/week/month *t+1* using information available up to and including day/week/month *t* and $\sigma_{realized,t}$ is the actual *t+1* volatility.

[6] The Google-enhanced model is less likely to work for weekly data than for daily- or monthly data since no Google search variable is significant at the 1% level for weekly data.

The forecast evaluations are presented in Table 8 where the Google-enhanced forecasting model, that acknowledges the relationship between search volumes on Google for the word “bitcoin”, is compared to the two benchmark models for the three different forecasting horizons. Like the earlier in-sample results the results for the weekly data (weekly forecasting horizon) deviates from the daily- and monthly results. For the latter two data sets the Google-enhanced model does indeed do a better job in predicting future Bitcoin volatility. For the weekly forecasts, however, the results are worse than for the two benchmark models. As mentioned above, this is not surprising considering the disappointing OLS-regression results for weekly data in Section B.

Regardless of forecasting horizon, and for every loss function, the most naïve forecasting model, i.e. the random walk-model, demonstrates a worse performance than the slightly more elaborate AR-model which tells us that past volatility changes have predictive power when it comes to forecasting Bitcoin volatility.

For daily and monthly forecasting horizons the enhanced forecasting model dominates the two benchmark models for every loss functions except in one single case, the RMSE for daily forecasts. This inconsistency might be linked to the particular features of the Bitcoin market. While the RMSE is widely used in evaluating volatility forecasts, it has some serious drawbacks that are particularly problematic in our case (where the Bitcoin volatility both is very high and very volatile). By construction, for the RMSE, single outliers (large forecast errors) increase the loss function significantly. This could be a problem in cases where one large error is not considered more troublesome than a sum of small errors. If that is the case, one possible solution is to rely on other loss functions, such as the R^2LOG. The R^2LOG also assigns higher weighting to large errors, but less so than the *RMSE*. The *PSE,* in turn, focuses on percentage errors and hence controls for the fact that it is harder to be accurate, in an absolute sense, when forecasting in high-volatility regimes. The same goes for the *QL* loss function. Compared to e.g. the *RMSE*, the *QL* loss function therefore makes it easier to compare forecasting ability across volatility regimes.

Overall, though, internet activity seems to be relevant for the behavior in the Bitcoin market and for anyone who wants to predict volatilities in the Bitcoin market it could pay off to acknowledge search pressure on search engines like Google. However, while this is likely to improve predictions, the improvement is somewhat limited and, as we have shown, also depends on both the forecasting horizon and on how the forecast accuracy is evaluated.

V. Conclusion

In this paper we look at the volatility in the Bitcoin market and how this volatility is related to the volatility in other relevant markets as well as to various market-wide risk indicators. We also investigate whether the volatility in the Bitcoin market is explained by retail investor-driven internet search volumes. The time period is 2011 to 2017 and daily, weekly as well as monthly data is employed. We contribute to the literature in several ways and our results could be of significant practical importance if the Bitcoin market continues to grow at the current speed. Our main finding, based on correlations, OLS-regressions and VAR-analysis, is a fairly strong positive link between Bitcoin volatility and search pressures on Bitcoin-related words on Google, particularly for the search word “bitcoin”. Other than that the only (somewhat) significant “driver” of Bitcoin volatility is the volatility in the USD currency market. We further show, using several different loss functions, that Google search activity can be used to make improved predictions of Bitcoin volatility. Overall, internet activity seems to be relevant for the behavior in the Bitcoin market and for anyone who wants to explain, understand or predict volatilities in the Bitcoin market it could pay off to acknowledge search pressure on search engines like Google.

Moreover, the significant link between Google search volumes and Bitcoin volatility points at retail investors, rather than large institutional investors, being major drivers of Bitcoin volatility.

REFERENCES

Antonopoulos, A. M. (2014). Mastering Bitcoin, O'Reilly Media Inc., Sebastopol, CA, US.

Baek, C., Elbeck, M. (2015). Bitcoins as an Investment or Speculative Vehicle? A First Look, Applied Economics Letters 22 (1), 30-34.

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring Economic Policy Uncertainty. The Quarterly Journal of Economics, 131(4), 1593-1636.

Balcilar, M., Bouri, E., Gupta, R., & Roubaud, D. (2017). Can Volume Predict Bitcoin Returns and Volatility? A Quantiles-Based Approach. Economic Modelling, 64, 74-81

Baur, D., Dimpfl, T. (2017). Realized Bitcoin Volatility, working paper, Available at SSRN: https://ssrn.com/abstract=2949754

Blockchain.com. (2018). Charts available at: https://blockchain.info

Bloomberg.com. (2018). Available at: https://www.bloomberg.com/news/articles/2017-12-15/turning-coal-into-Bitcoin-dirty-secret-of-2017-s-hottest-market

Bollerslev, T., Engle, R.F. and Nelson, D, (1994). Handbook of Econometrics, ed. by Engle, R.F. and McFadden, D.L., vol IV, pp. 2961-3038.

Bouri, E., Molnár, P., Azzi, G., Roubaud, D., & Hagfors, L. I. (2017). On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier? Finance Research Letters, 20, 192-198.

Bouri, E., Azzi, G., & Dyhrberg, A. H. (2016). On the return-volatility relationship in the Bitcoin market around the price crash of 2013. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2869855

Brownlees, C., Engle, R. and Kelly, B. A practical guide to volatility forecasting through calm and storm. Journal of Risk, 2011, 14(2), pp. 1–20.

Ciaian, P., Rajcaniova, M., & Kancs, D. A. (2016). The Economics of Bitcoin Price Formation. Applied Economics, 48(19), 1799-1815.

Coinmarketcap.com. (2018). Charts available at: https://coinmarketcap.com

Dyhrberg, A. H. (2016). Bitcoin, Gold and the Dollar–A GARCH Volatility Analysis. Finance Research Letters, 16, 85-92.

Financial Times (2017a). Chicago Derivatives Exchanges do Battle to Dominate Bitcoin Futures Trading, Financial Times, Dec 4, 2017, Available at: https://www.ft.com/content/e535b17c-d8ea-11e7-a039-c64b1c09b482

Financial Times (2017b). ETF Providers Hope Launch of Bitcoin Derivatives Will Help Swap Regulators to Approve Trackers, Financial Times, Dec 17, 2017, Available at: https://www.ft.com/content/28bcfc66-e1d8-11e7-a8a4-0a1e63a52f9c

Financial Times (2017c). Big Investors Yet to Invest in Bitcoin, Financial Times, Oct 20, 2017, Available at: https://www.ft.com/content/4c700f9a-b267-11e7-aa26-bb002965bce8

Financial Times (2018). Bitcoin Whales Control a Third of Market with Holdings of $37.5bn, Financial Times, June 9, 2018, Available at: https://www.ft.com/content/c4b68aec-6b26-11e8-8cf3-0c230fa67aec

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

Kaminski, J. (2014). Nowcasting the bitcoin market with twitter signals. arXiv preprint arXiv:1406.7577

Kristoufek, L. (2013). Bitcoin meets Google Trends and Wikipedia: Quantifying the Relationship Between Phenomena of the Internet Era. Scientific Reports 3 (3415), 1–7.

Kristoufek, L. (2015). What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. PloS one, 10(4), e0123923.

MotleyFool. (2017). Available at: https://www.fool.com/investing/2017/12/30/Bitcoin-investors-are-irrational-and-this-figure-p.aspx

Pesaran, H. H., & Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models. Economics letters, 58(1), 17-29.

Saldías, M. (2013). Systemic risk analysis using forward-looking distance-to-default series. Journal of Financial Stability, 9(4), 498-517.

Urquhart, A. (2018). What Causes the Attention of Bitcoin? Available at SSRN: https://ssrn.com/abstract=3097153

Wong, J. I. (2014). Bitmex to Launch Bitcoin 'Fear' Index, Coindesk.com, Available at: https://www.coindesk.com/bitmex-launch-Bitcoin-fear-index/

Yelowitz, A., Wilson, M. (2015). Characteristics of Bitcoin Users: an Analysis of Google Search Data, Applied Economics Letters, 22 (13), 1030–1036.

Yermack, D. (2015). Is Bitcoin a real currency? An economic appraisal. In Handbook of digital currency (pp. 31-43).

Table 1 Summary statistics for variables. Variables are expressed in log first differences and on a daily basis, covering the period December 2016 to June 2017. The variables ending in *VOL* are time series of volatilities, calculated as squared price changes for the daily frequency. *EPU* and *SYS* are the Economic Policy Uncertainty Index (Baker et al. (2016)) and Systemic Risk Indicator (provided by the Cleveland Federal Reserve). The following nine variables starting with *G* are Google search strings collected from Google Trends.

Variable	Obs	Mean	Std. Dev.	Min	Max	Skew	Kurt
BTCVOL	128	0.0005	1.7164	-5.7149	4.5757	-0.2206	3.6811
GOLDVOL	122	0.0313	1.8863	-5.1341	5.1704	-0.2317	3.7972
USDVOL	125	0.0255	1.2766	-3.0245	4.0038	0.2695	3.4247
S&P500VOL	113	-0.0249	1.4702	-4.6913	3.3819	-0.0791	3.1996
EPU	128	0.0055	0.5182	-1.1691	1.4985	0.2022	3.4530
SYS	128	0.0009	0.0204	-0.0982	0.0696	-0.5967	7.5728
GBTC	128	0.0075	0.1377	-0.4182	0.4626	0.1772	4.5887
GVIX	128	-0.0012	0.1504	-0.4895	0.5293	0.4846	5.5071
GCRISIS	128	0.0001	0.0842	-0.2106	0.2201	0.2137	2.9222
GCYBER	126	0.0000	0.4629	-0.9555	4.3820	6.5624	65.5548
GGOLD	128	-0.0030	0.0602	-0.2458	0.2097	-0.3408	5.5931
GINTE	128	-0.0022	0.4235	-1.3218	1.0986	-0.2342	3.2015
GINFL	128	0.0009	0.1081	-0.2919	0.3221	0.1343	2.9710
GCRASH	128	-0.0015	0.2194	-0.6931	0.6690	-0.1547	3.9575
GWAR	128	-0.0006	0.0610	-0.3011	0.4155	1.6317	22.5964

Table 2 Summary statistics for variables. Variables are expressed in log first differences and on a weekly basis, covering the period June 2012 to June 2017. The variables ending in *VOL* are time series of volatilities, calculated as ordinary weekly sample standard deviations based on the daily volatility. *EPU* and *SYS* are the Economic Policy Uncertainty Index (Baker et al. (2016)) and Systemic Risk Indicator (provided by the Cleveland Federal Reserve). The following nine variables starting with *G* are Google search strings collected from Google Trends.

Variable	Obs	Mean	Std. Dev.	Min	Max	Skew	Kurt
BTCVOL	260	0.0005	0.8691	-3.3334	2.6025	-0.0075	3.8458
GOLDVOL	260	0.0009	0.5989	-1.4477	1.7277	-0.1119	2.6917
USDVOL	260	-0.0024	0.5654	-1.7049	1.9387	-0.0525	3.0123
S&P500VOL	260	-0.0061	0.7092	-2.2248	2.7854	0.1037	4.0483
EPU	260	-0.0026	0.3269	-1.0729	0.9126	-0.0600	4.0095
SYS	260	0.0014	0.0596	-0.3717	0.1683	-1.4568	10.0362
GBTC	260	0.0129	0.2085	-0.6931	0.9163	0.7069	6.3837
GVIX	260	0.0010	0.1356	-0.5476	0.6733	0.2247	6.9646
GCRISIS	260	-0.0005	0.0939	-0.4636	0.5447	0.1392	10.2082
GCYBER	260	0.0053	0.5591	-2.6593	3.6109	0.9436	11.9470
GGOLD	260	0.0006	0.1697	-0.6539	1.5606	3.7104	34.2488
GINTE	260	0.0040	0.3394	-1.6607	1.0986	-0.5290	6.0495
GINFL	260	0.0000	0.0901	-0.4177	0.3264	-0.8421	8.2030
GCRASH	260	0.0027	0.3679	-1.8068	1.9021	0.5598	11.5323
GWAR	260	-0.0003	0.0675	-0.2432	0.2423	0.3899	5.5501

Table 3 Summary statistics for variables. Variables are expressed in log first differences and on a monthly basis, covering the period August 2011 to June 2017. The variables ending in *VOL* are time series of volatilities, calculated as ordinary monthly sample standard deviations based on the daily volatility. *EPU* and *SYS* are the Economic Policy Uncertainty Index (Baker et al. (2016)) and Systemic Risk Indicator (provided by the Cleveland Federal Reserve). The following nine variables starting with *G* are Google search strings collected from Google Trends.

Variable	Obs	Mean	Std. Dev.	Min	Max	Skewness	Kurtosis
BTCVOL	70	-0.0066	0.6941	-1.3018	1.3882	-0.0126	2.0721
GOLDVOL	70	-0.0192	0.3792	-0.6984	1.8478	1.6810	9.9613
USDVOL	70	-0.0011	0.3060	-0.7137	0.9414	0.3043	3.2740
S&P500VOL	70	-0.0180	0.4813	-1.5213	1.3273	-0.1335	3.7557
EPU	70	-0.0106	0.2512	-0.8241	0.7890	0.0944	4.7893
SYS	70	0.0115	0.0887	-0.2423	0.2062	-0.5105	3.4490
GBTC	70	0.0402	0.3634	-0.9808	1.1939	0.9072	5.5467
GVIX	70	-0.0055	0.1614	-0.5108	0.4220	0.2690	4.3717
GCRISIS	70	-0.0029	0.1241	-0.2829	0.4447	0.4487	4.6414
GCYBER	70	0.0374	0.5807	-1.2993	2.8134	1.8574	10.3520
GGOLD	70	-0.0089	0.1994	-0.4620	1.0217	2.2381	12.8415
GINTE	70	0.0048	0.2872	-0.9163	0.9295	-0.1650	4.7834
GINFL	70	-0.0029	0.0892	-0.1780	0.2136	-0.1082	2.6579
GCRASH	70	-0.0118	0.3759	-1.0046	1.0846	0.0047	4.3101
GWAR	70	-0.0024	0.1062	-0.4943	0.2107	-1.2171	8.2919

Table 4 Correlation matrix expressing the pairwise correlation between variables over the sample period December 2016 to June 2017 for the daily frequency. The number of observations is 125. Correlation is based on variables being expressed in log first differences. The significance level of the correlation coefficient is indicated by * (10%), ** (5%) and *** (1%). The variables ending in *VOL* are time series of volatilities, calculated as squared price changes for the daily frequency. *EPU* and *SYS* are the Economic Policy Uncertainty Index (Baker et al. (2016)) and Systemic Risk Indicator (provided by the Cleveland Federal Reserve). The following nine variables starting with *G* are Google search strings collected from Google Trends.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
BTCVOL	1.00														
GOLDVOL	0.15*	1.00													
USDVOL	0.11	0.23**	1.00												
S&P500VOL	-0.08	-0.02	-0.07	1.00											
EPU	-0.11	-0.14	-0.04	-0.03	1.00										
SYS	-0.08	-0.09	-0.02	0.00	0.18**	1.00									
GBTC	0.21**	0.15	0.19**	0.01	0.00	0.10	1.00								
GVIX	0.21**	0.18**	0.17*	0.13	-0.17*	-0.20**	-0.08	1.00							
GCRISIS	0.18**	0.15	0.11	0.01	-0.20**	-0.13	0.02	0.42***	1.00						
GCYBER	0.01	-0.01	0.28***	-0.07	0.13	-0.05	0.13	-0.18**	-0.11	1.00					
GGOLD	-0.21**	0.07	0.14	0.09	-0.06	-0.01	0.14	0.07	0.01	-0.07	1.00				
GINTE	0.23**	-0.04	0.03	0.09	0.00	-0.04	0.07	0.17*	0.26***	-0.13	0.00	1.00			
GINFL	0.25***	0.27***	0.29***	0.03	-0.14	-0.06	-0.03	0.38***	0.65***	-0.06	-0.14	0.19**	1.00		
GCRASH	0.35***	0.29***	0.19**	0.10	-0.15*	-0.15*	0.02	0.54***	0.47***	-0.11	0.12	0.13	0.45***	1.00	
GWAR	-0.20**	0.06	0.03	0.03	0.06	0.00	-0.02	-0.01	0.08	-0.06	0.30***	0.02	-0.09	0.01	1.00

Table 5 Correlation matrix expressing the pairwise correlation between variables over the sample period June 2012 to June 2017 for the weekly frequency. The number of observations is 260. Correlation is based on variables being expressed in log first differences. The significance level of the correlation coefficient is indicated by * (10%), ** (5%) and *** (1%). The variables ending in *VOL* are time series of volatilities, calculated as ordinary weekly sample standard deviations based on the daily volatility. *EPU* and *SYS* are the Economic Policy Uncertainty Index (Baker et al. (2016)) and Systemic Risk Indicator (provided by the Cleveland Federal Reserve). The following nine variables starting with *G* are Google search strings collected from Google Trends.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
BTCVOL	1.00														
GOLDVOL	0.12*	1.00													
USDVOL	0.12	0.30***	1.00												
S&P500VOL	-0.08	0.25***	0.33***	1.00											
EPU	-0.09	0.12*	0.29***	0.26***	1.00										
SYS	0.03	-0.13**	-0.04	-0.31***	-0.20***	1.00									
GBTC	-0.11*	-0.03	-0.04	0.00	0.09	0.04	1.00								
GVIX	0.16***	-0.06	0.03	-0.10	0.02	0.13**	0.03	1.00							
GCRISIS	-0.08	-0.01	0.03	-0.01	0.05	0.03	0.09	0.21***	1.00						
GCYBER	0.08	-0.13**	-0.01	-0.09	-0.08	0.12*	0.07	0.09	0.07	1.00					
GGOLD	0.16**	0.06	-0.07	-0.08	-0.12*	0.10	0.00	0.23***	0.01	-0.02	1.00				
GINTE	-0.14**	-0.11*	-0.07	-0.06	0.09	-0.05	0.05	0.14**	0.16**	-0.04	-0.01	1.00			
GINFL	-0.04	-0.01	0.00	-0.06	0.05	0.07	0.08	0.21***	0.58***	0.11*	0.04	0.47***	1.00		
GCRASH	0.19***	0.05	0.03	-0.07	-0.07	0.10	0.13*	0.54***	0.34***	0.16***	0.23***	0.13**	0.40***	1.00	
GWAR	-0.12**	-0.11*	-0.07	0.03	0.02	0.03	0.04	0.01	0.23***	-0.02	0.11*	0.10	0.18***	0.08	1.00

Table 6 Correlation matrix expressing the pairwise correlation between variables over the sample period August 2011 to June 2017 for the monthly frequency. Number of observations is 70. Correlation is based on variables being expressed in log first differences. The significance level of the correlation coefficient is indicated by * (10%), ** (5%) and *** (1%). The variables ending in *VOL* are time series of volatilities, calculated as ordinary monthly sample standard deviations based on the daily volatility. *EPU* and *SYS* are the Economic Policy Uncertainty Index (Baker et al. 2016) and Systemic Risk Indicator (provided by the Cleveland Federal Reserve). The following nine variables starting with *G* are Google search strings collected from Google Trends

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
BTCVOL	1.00														
GOLDVOL	0.07	1.00													
USDVOL	0.33***	0.15	1.00												
S&P500VOL	0.06	0.38***	0.27**	1.00											
EPU	0.03	0.27**	0.30**	0.19	1.00										
SYS	-0.05	-0.36***	-0.08	-0.50***	-0.31***	1.00									
GBTC	0.56***	0.13	0.05	0.03	-0.14	0.14	1.00								
GVIX	0.15	0.20*	0.24**	0.49***	0.33***	-0.23*	0.09	1.00							
GCRISIS	-0.04	0.02	-0.22*	0.00	-0.19	-0.16	-0.04	-0.07	1.00						
GCYBER	-0.03	-0.09	0.09	-0.18	-0.06	0.22*	0.21*	-0.02	0.21*	1.00					
GGOLD	0.31**	0.58***	0.03	0.07	0.05	-0.07	0.35***	0.04	-0.03	-0.02	1.00				
GINTE	-0.12	0.11	0.01	0.16	-0.18	0.05	-0.05	0.12	0.17	-0.07	0.07	1.00			
GINFL	0.08	0.10	0.05	0.03	-0.17	-0.02	0.05	-0.05	0.56***	0.13	0.01	0.30**	1.00		
GCRASH	0.24*	0.13	0.09	0.39***	0.13	-0.10	0.14	0.62***	0.25**	-0.01	0.21*	0.14	0.39***	1.00	
GWAR	-0.21*	-0.10	-0.29**	-0.08	-0.28**	0.00	-0.10	-0.32***	0.29*	-0.01	-0.17	0.16	0.38***	0.00	1.00

Table 7 OLS regression results. Two OLS models are estimated for each frequency (daily, weekly and monthly). One (even numbered) including one period lagged Bitcoin volatility, and one (odd numbers) without. All variables are expressed in logged fist differences and the significance level is indicated by * (10%), ** (5%) and *** (1%). The dependent variable is the Bitcoin volatility. The variables ending in *VOL* are time series of volatilities. *EPU* and *SYS* are the Economic Policy Uncertainty Index (Baker et al. 2016) and Systemic Risk Indicator (provided by the Cleveland Federal Reserve). The following nine variables starting with *G* are Google search strings collected from Google Trends.

	DAILY		WEEKLY		MONTHLY	
VARIABLE	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
GOLDVOL	0.0033 *(0.0940)*	-0.0485 *(0.0826)*	0.1166 *(0.0941)*	0.0817 *(0.0853)*	-0.3216 *(0.2642)*	-0.3238 *(0.2375)*
USDVOL	0.0069 *(0.1387)*	0.06157 *(0.1215)*	0.2043** *(0.1028)*	0.1755* *(0.0929)*	0.8044*** *(0.2535)*	0.6747*** *(0.2338)*
S&P500VOL	-0.1125 *(0.1055)*	-0.1277 *(0.0922)*	-0.1437* *(0.0829)*	-0.0553 *(0.0757)*	-0.2192 *(0.2013)*	-0.2168 *(0.1842)*
EPU	-0.3022 *(0.3194)*	-0.1665 *0.2800)*	-0.1678 *(0.1728)*	-0.2676* *(0.1565)*	-0.1207 *(0.3210)*	-0.2068 *(0.2871)*
SYS	-1.9709 *(7.4448)*	-1.3419 *(6.5020)*	-0.6929 *(0.9341)*	-1.2321 *(0.8461)*	-1.3333 *(0.9909)*	-1.4741* *(0.8749)*
GBTC	2.6364** *(1.2968)*	3.0099*** *(1.1344)*	-0.4267* *(0.2512)*	-0.2840 *(0.2275)*	1.0105*** *(0.2056)*	0.9912*** *(0.1825)*
GVIX	0.4762 *(1.3090)*	0.5228 *(1.1431)*	0.5316 *(0.4636)*	0.5238 *(0.4187)*	-0.0825 *(0.6921)*	-0.0043 *(0.6294)*
GCRISIS	-3.0805 *(2.6404)*	-1.3695 *(2.3261)*	-1.1842* *(0.7004)*	-0.8726 *(0.6347)*	0.2324 *(0.7173)*	0.2598 *(0.6329)*
GCYBER	0.0097 *(0.3479)*	-0.0373 *(0.3039)*	0.0832 *(0.0952)*	0.1201 *(0.0861)*	-0.2348* *(0.125)*	-0.2863** *(0.1119)*
GGOLD	-6.2357** *(2.7787)*	-2.5973 *(2.5131)*	0.5037 *(0.3212)*	0.6265** *(0.2907)*	0.5835 *(0.4995)*	0.5841 *(0.4455)*
GINTE	0.9365** *(0.4018)*	0.8423** *(0.3513)*	-0.3447* *(0.1783)*	-0.2826* *(0.1613)*	-0.2559 *(0.2585)*	-0.1711 *(0.2356)*
GINFL	1.3667 *(2.2308)*	-0.0833 *(1.9654)*	0.2908 *(0.8271)*	0.3902 *(0.7473)*	0.4261 *(1.1398)*	0.6511 *(1.0121)*
GCRASH	2.5214*** *(0.9598*	1.9897** *(0.8436)*	0.3922** *(0.1843)*	0.2330 *(0.1677)*	0.2959 *(0.31457)*	0.2815 *(0.2814)*
GWAR	-3.5145 *(2.5903)*	-3.6838 *(2.2622)*	-1.0303 *(0.7976)*	-0.4339 *(0.7248)*	-0.5891 *(0.7576)*	-0.9814 *(0.6943)*
BTCVOL(-1)		-0.4564*** *(0.0821)*		-0.4221*** *(0.0558)*		-0.3546*** *(0.0848)*
C	-0.0593 *(0.1506)*	-0.0437 *(0.1316)*	0.0041 *(0.0512)*	0.0060 *(0.0463)*	-0.0238 *(0.0654)*	-0.0290 *(0.0586)*
R2	0.2828	0.4588	0.1519	0.3140	0.5237	0.6428
Adjusted R2	0.1782	0.3734	0.1034	0.2717	0.4024	0.5417
n	111	110	260	259	70	69

Table 8 Three constructed bitcoin volatility forecast models are evaluated by means of four loss functions: the root mean square error (RMSE). the percentage squared error (PSE), the quasi likelihood (QL) and the R^2LOG loss functions. The three forecast models are a random walk model, where the predicted next-period volatility is assumed to be unchanged from the current-period volatility, an AR(1) model (the predicted next-period volatility is assumed to be related to the current-period volatility through the auto-regressive relationship of order one), and our Google-enhanced model, where an additional predictor, apart from the lagged bitcoin volatility, is the GBTC.

DAILY

	RMSE	PSE	QL	R^2LOG
Random walk	6013.5825	661.2124	9.6680	1.3071
AR	**4747.3322**	140.9577	4.1145	0.9040
Google	13451.6238	**116.6334**	**4.0754**	**0.8536**

WEEKLY

	RMSE	PSE	QL	R^2LOG
Random walk	55893.7095	672.6074	5.8122	0.6172
AR	**55791.8137**	**145.5440**	**3.4649**	**0.4973**
Google	64099.7098	851.8191	6.2606	0.6420

MONTHLY

	RMSE	PSE	QL	R^2LOG
Random walk	8308.4612	7.8767	1.4418	0.4445
AR	6325.1436	6.1275	1.1530	0.3065
Google	**4980.1099**	**3.6634**	**0.7220**	**0.1836**

Figure 1a Impulse response functions of bitcoin volatility for the daily frequency, based on data during December 2016 to June 2017. One period hence translates into one (trading) day, which is the x-axis. The y-axis corresponds to the logged first difference unit of the variable in question. The impulse corresponds to a one standard deviation shock to the residuals of the variable that is being hit by a shock (here called innovation). The line corresponds to the response, and the dotted lines indicate a 95 % confidence interval of the response. All of the variables are transformed and expressed in logged first differences. Interpretation of the impulse response goes as follows (top left picture as an example): if daily Googles searches of Bitcoin are being shocked by one standard deviation today, then daily Bitcoin volatility will increase by approximately 0.6 per cent (expressed in logged first difference) in the first period (day), i.e. tomorrow. As we can see from this example, the increase is statistically significant at the 5% significance level as both dotted lines lie above the x-axis. Impulse responses are derived from bivariate VAR(p) models with a lag length p of 3-7 for all variables. Lag length choices are based on LR, FPE, AIC, SC and HQ criteria. As expected. the impulse response goes to zero as time passes; almost directly after period 10 in all of the below impulse response functions. In some figures, the x-axis reaches up to 20 periods. This is only for clarification purposes to ascertain that the impulse response actually does revert over time.

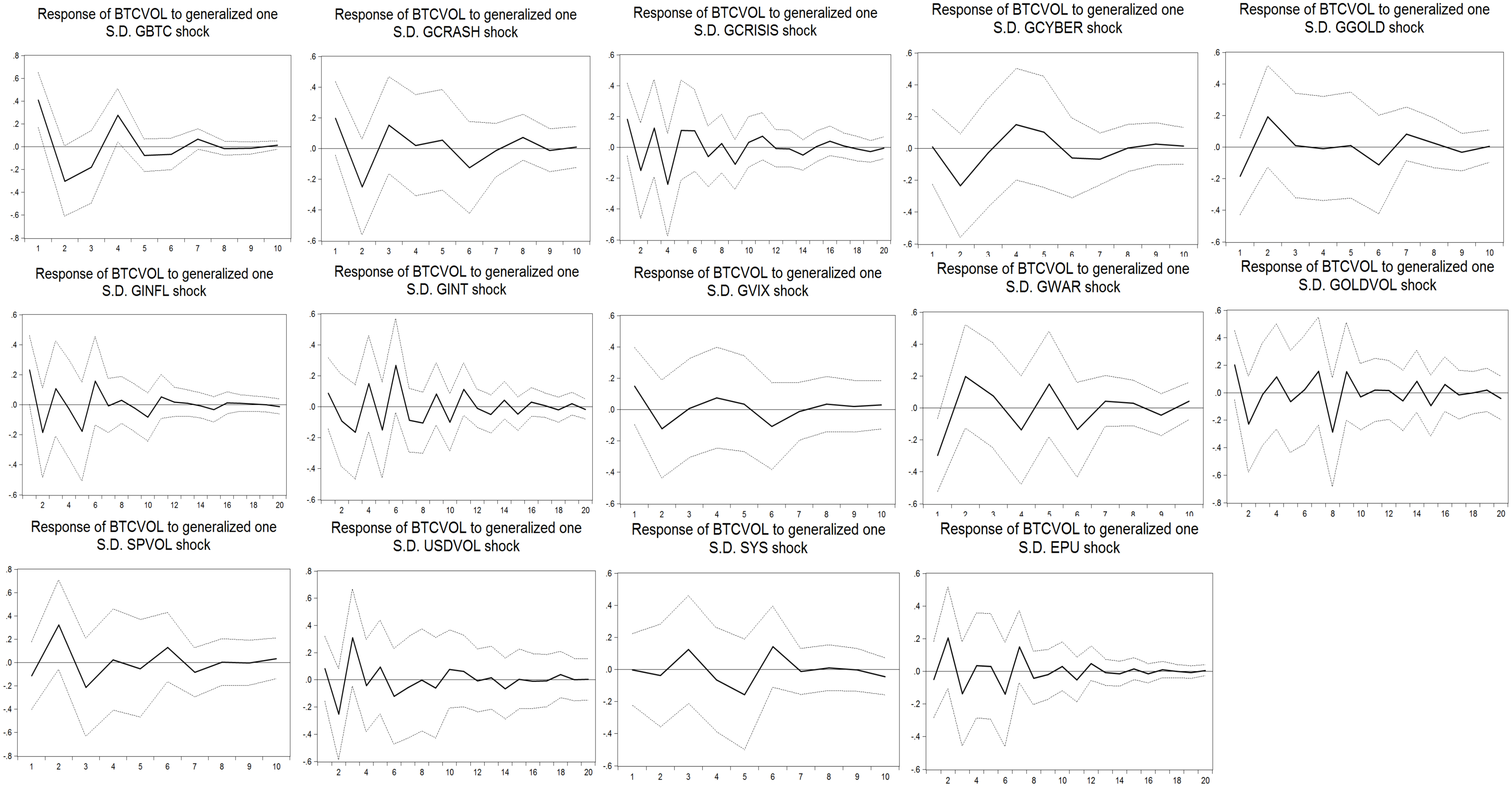

Figure 1b The corresponding accumulated impulse response of bitcoin volatility for shocks in all variables, based on figure 1a. As can be seen, in almost no cases is the accumulated response statistically significant. A significant response can only be seen in the Google variable BTC, where in period 1 and 4, the accumulated response at those times is statistically significant on the 5 % level. Dotted lines correspond to a 95% confidence interval.

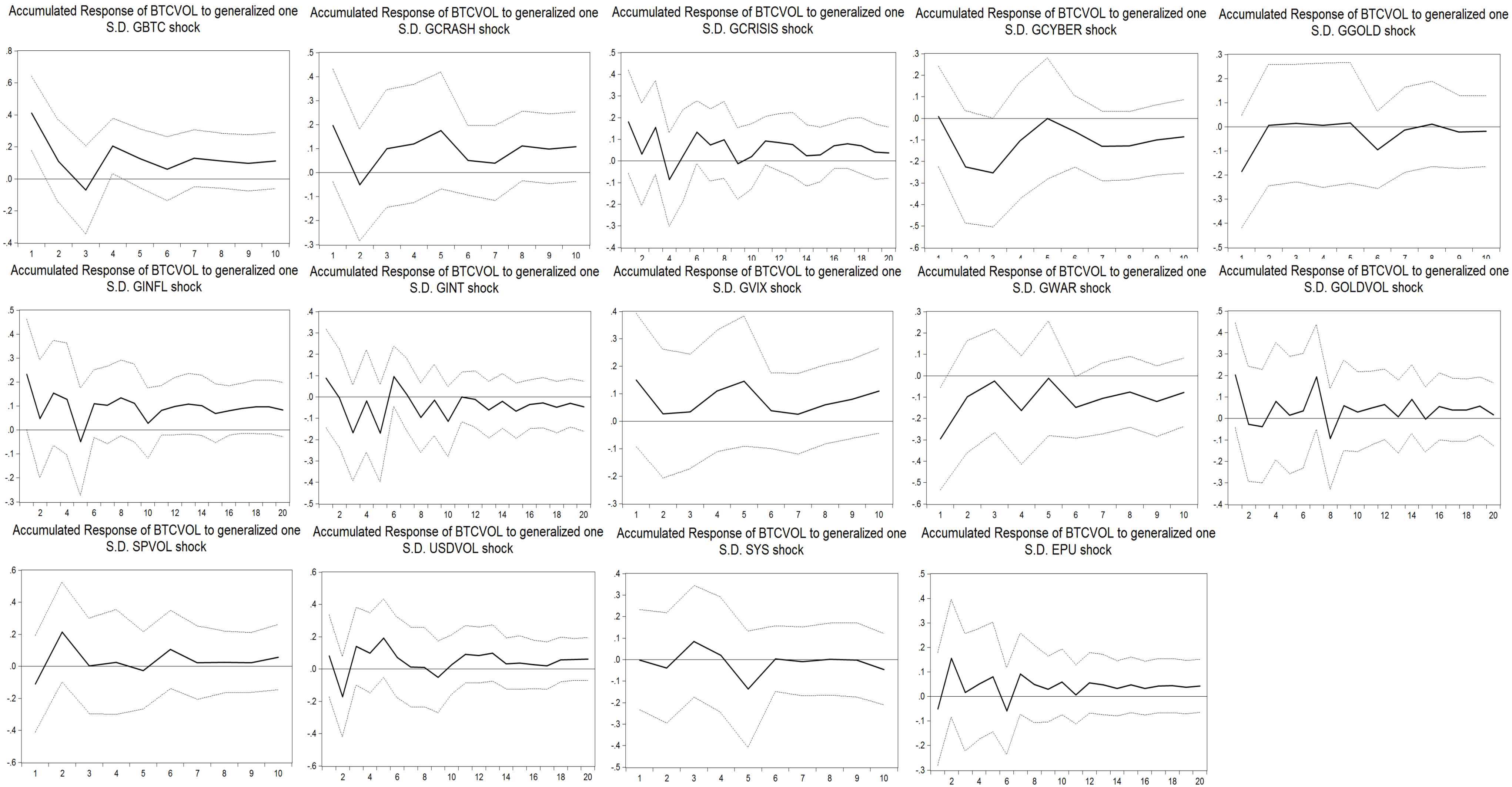

Figure 2a Impulse response functions of bitcoin volatility for the weekly frequency during the period June 2012 to June 2017. One period hence translates into one week (Mon – Fri). which is the x-axis. The y-axis corresponds to the logged first difference unit of the variable in question. The impulse corresponds to a one standard deviation shock to the residuals of the variable that is being hit by a shock (here called innovation). The line corresponds to the response, and the dotted lines indicate a 95 % confidence interval of the response. All of the variables are transformed and expressed in logged first differences. Interpretation of the impulse response goes as follows (top left picture as an example): if weekly Googles searches of Bitcoin are being shocked by one standard deviation, then weekly Bitcoin volatility will decrease by approximately 0.1 per cent (expressed in logged first difference) in the first period (week), and then increase by approximately 0.45 percent in the second period. As we can see from this example, both the decrease and increase seem to be statistically significant on a 5% significance level as both dotted lines lie below (period 1) and above (period 2) the x-axis. Impulse responses are derived from bivariate VAR(6) models. The lag length choice is based on LR, FPE, AIC, SC and HQ criteria. As expected, the impulse response goes to zero as time passes; almost directly after period 10 in all of the below impulse response functions. In some figures, the x-axis reaches up to 20 periods. This is only for clarification purposes to ascertain that the impulse response actually does revert over time.

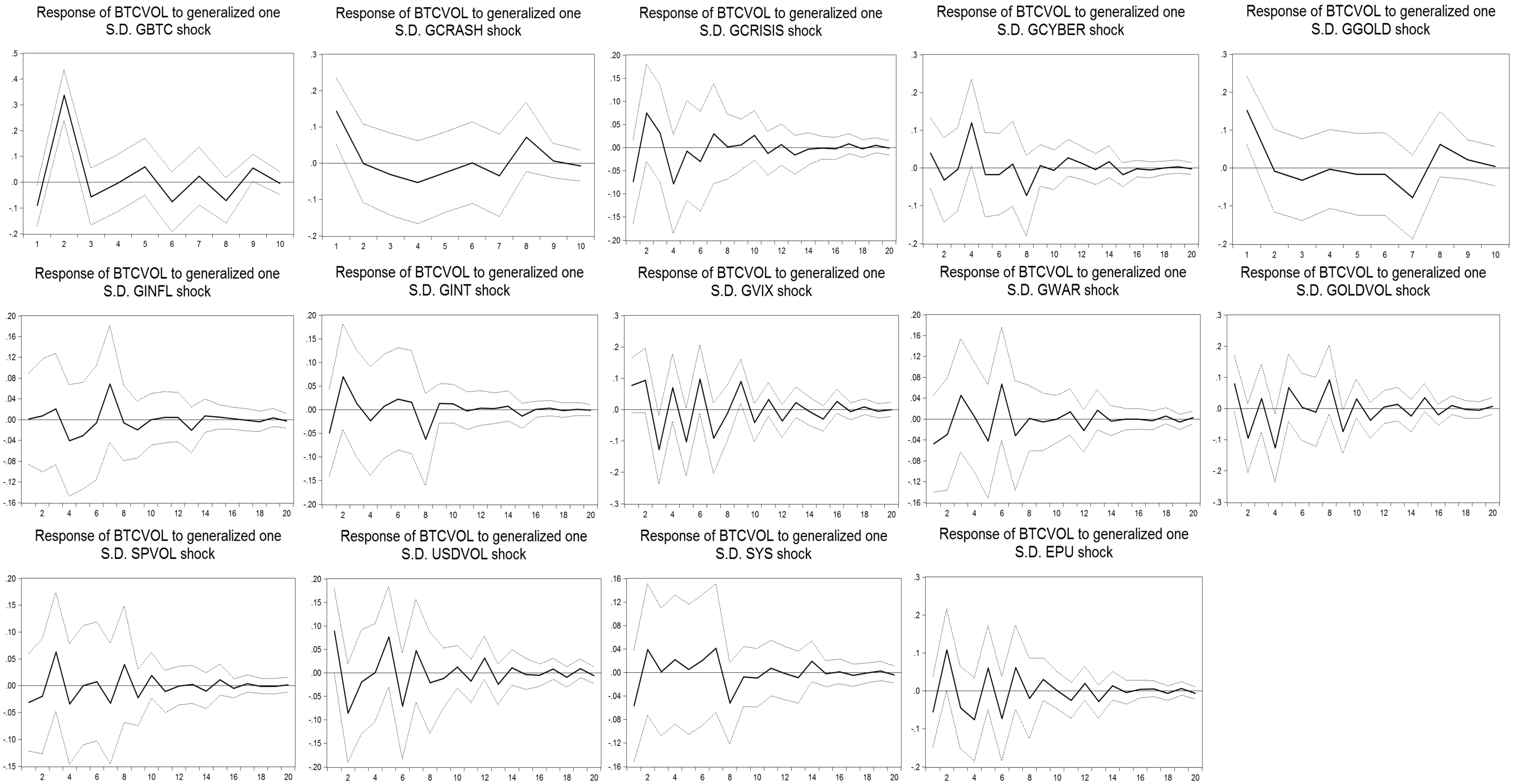

Figure 2b The corresponding accumulated impulse response of bitcoin volatility for shocks in all variables, based on figure 2a. Dotted lines correspond to a 95% confidence interval. As can be seen, shocks in GBTC, GCRASH, GCYBER, GGOLD and GVIX, almost all Google variables, generate a statistically significant accumulated response in Bitcoin volatility. One period (x-axis) corresponds to one week.

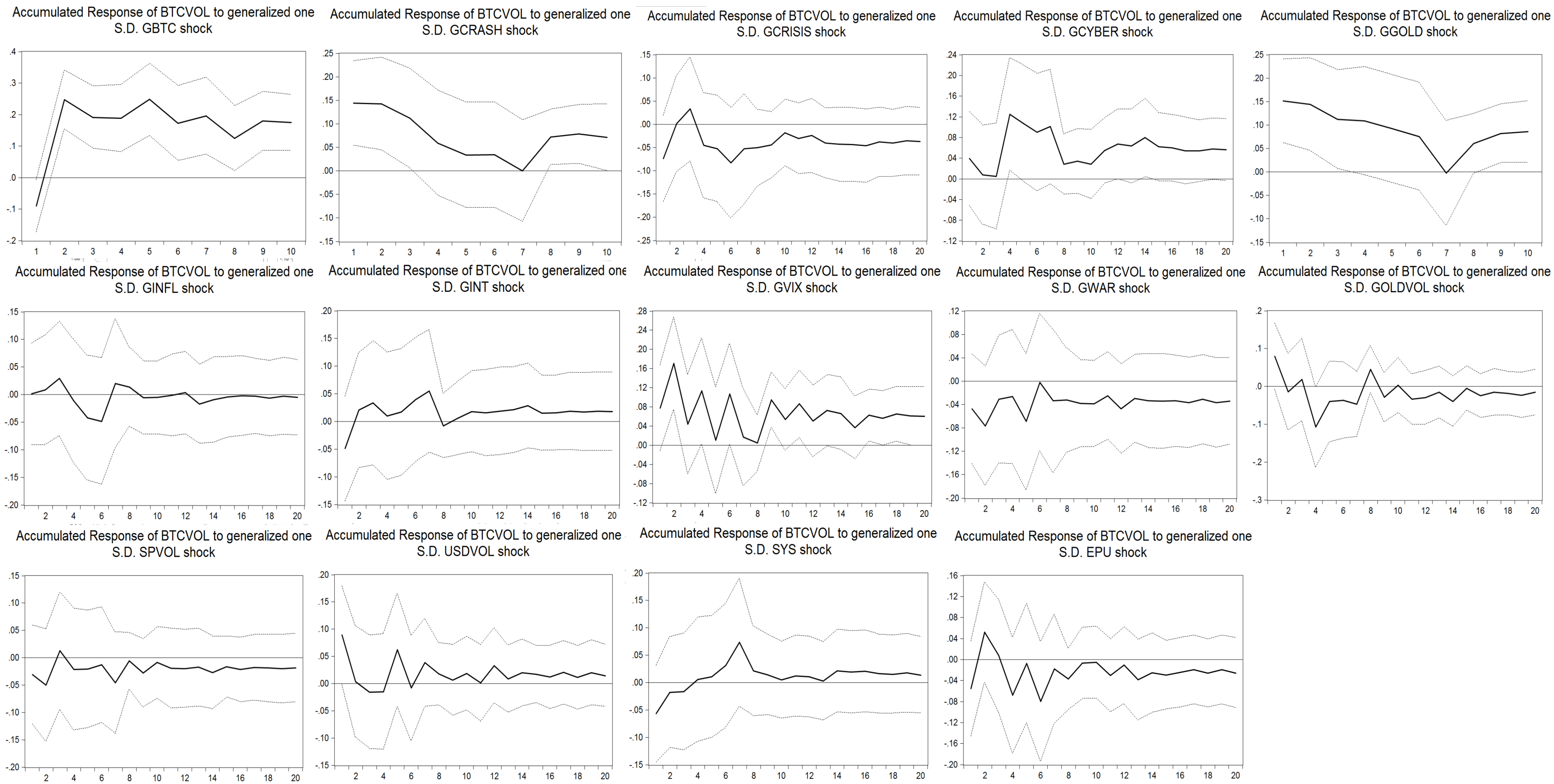

Figure 3a Impulse response functions of bitcoin volatility for the monthly frequency during the period August 2011 to June 2017. One period hence translates into one month, which is the x-axis. The y-axis corresponds to the logged first difference unit of the variable in question. The impulse corresponds to a one standard deviation shock to the residuals of the variable that is being hit by a shock (here called innovation). The line corresponds to the response, and the dotted lines indicate a 95 % confidence interval of the response. All of the variables are transformed and expressed in logged first differences. Interpretation of the impulse response goes as follows (top left picture as an example): if monthly Googles searches of Bitcoin are being shocked by one standard deviation, then monthly Bitcoin volatility will rise by approximately 0.3 per cent (expressed in logged first difference). As we can see from this example, the increase in volatility is also statistically significant on a 5% significance level as both dotted lines lie above the x-axis. Impulse responses are derived from bivariate VAR(p)-models with a lag length p (1 to 5) depending on the variable. Lag length choices are based on LR, FPE, AIC, SC and HQ criteria. As expected, the impulse response goes to zero as times passes; almost directly after period 10 in all of the below impulse response functions.

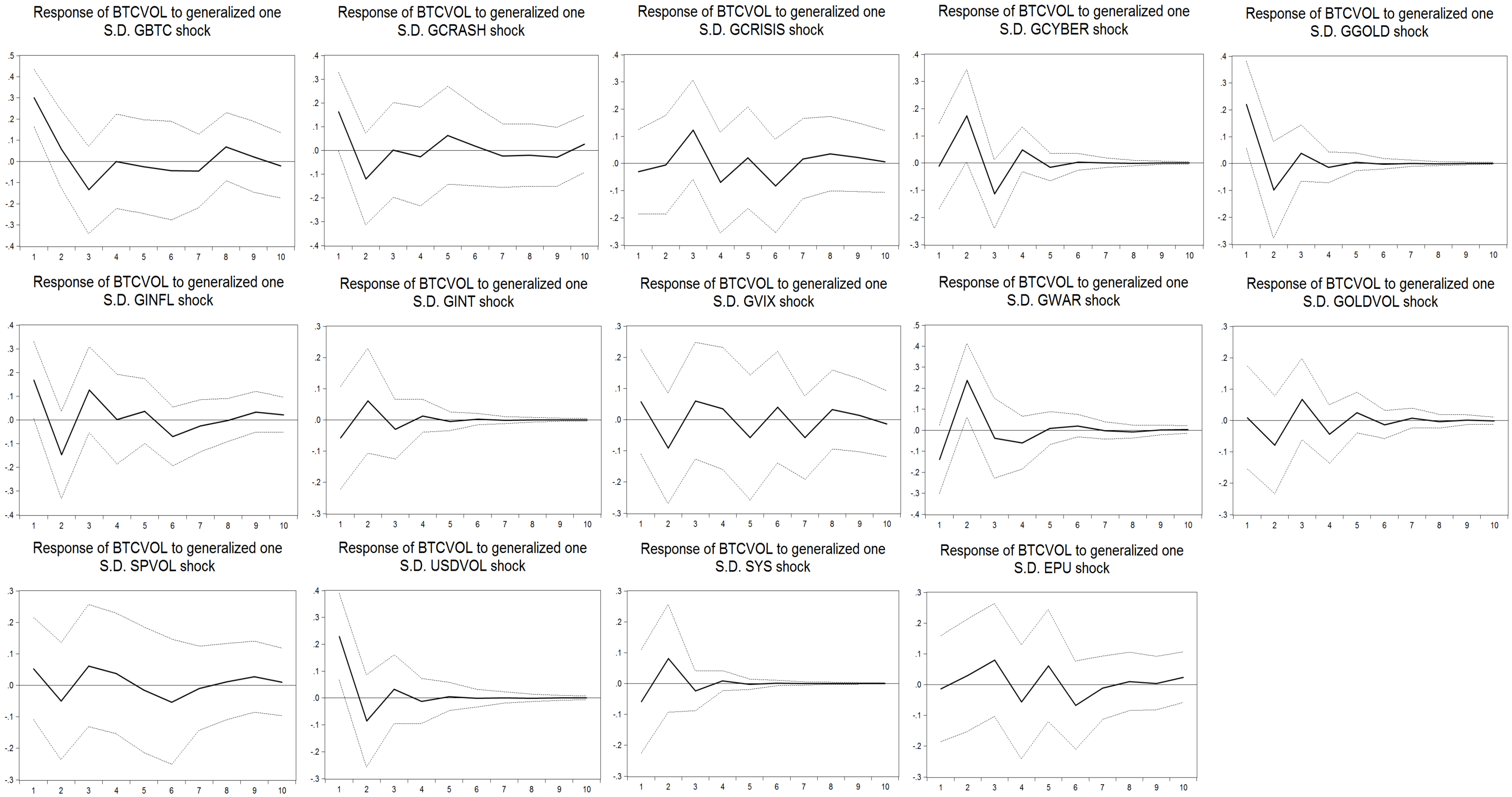

Figure 3b The corresponding accumulated impulse response of bitcoin volatility for shocks in all variables, based on figure 3a. Dotted lines correspond to a 95% confidence interval, and one period (x-axis) corresponds to one month. As can be seen from the figures, GBTC, GGOLD and USDVOL generate a statistically significant accumulated response in the Bitcoin volatility variable when they are shocked by one standard deviation. In most cases however, the accumulated response cannot be proven to be different from zero.

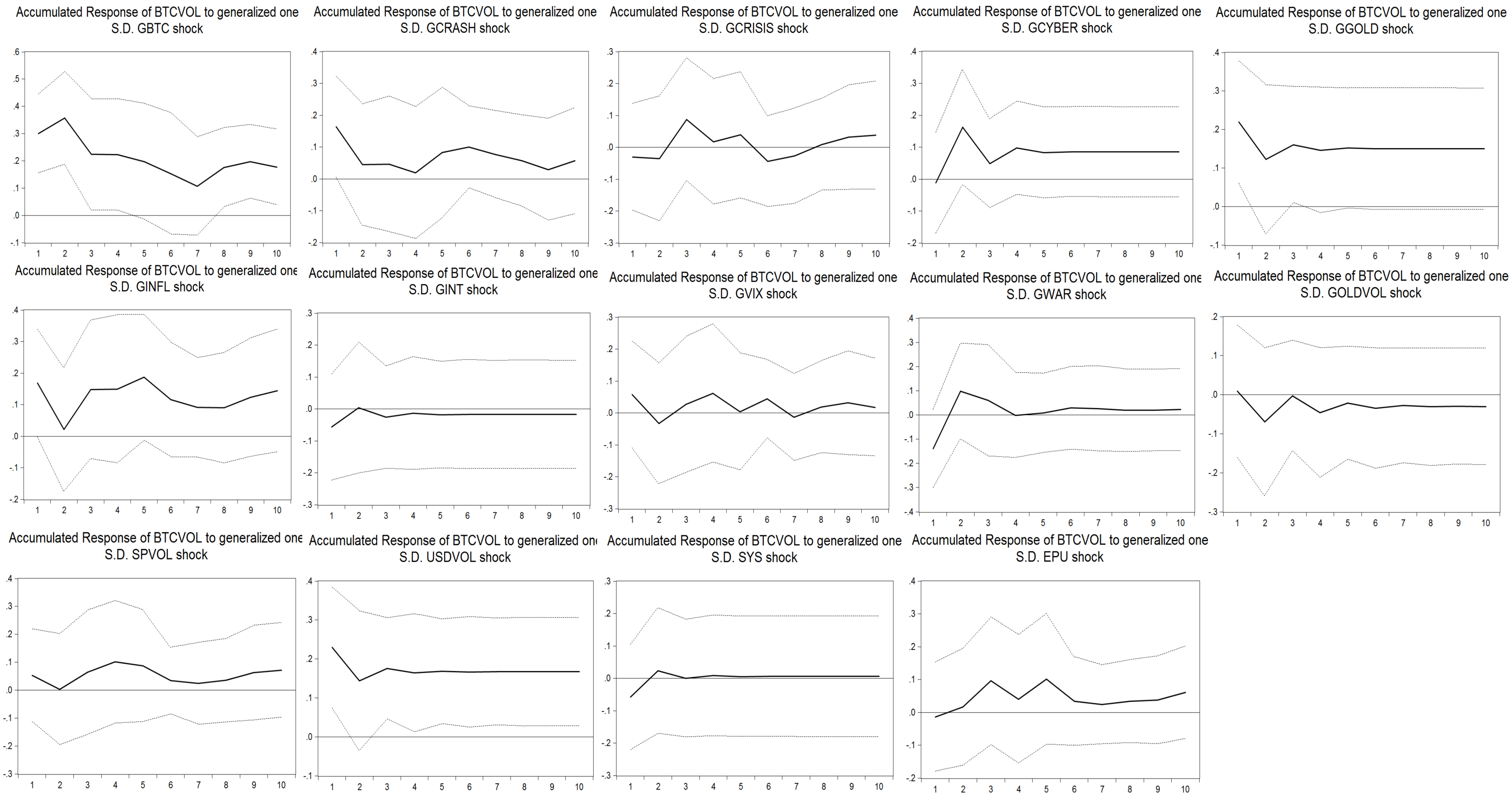

Figure 4 Impulse response and accumulated impulse functions (second row) of bitcoin volatility for one standard deviation shocks to the Google variables BTC, Crash, Gold and Interest rate, based on data during December 2016 to June 2017. The x-axis represents the time period, in this case 1 equals one day, the y-axis corresponds to the unit of measurement of the bitcoin volatility (log first difference). The impulse corresponds to a one standard deviation shock to the residuals of the variable that is being hit by a shock. The line corresponds to the response, and the dotted lines indicate a 95 % confidence interval of the response. Impulse responses are derived from a VAR(2) model with five variables; the bitcoin volatility and the Google variables BTC, Crash, Gold and Interest rate. The lag length choice is based on LR, FPE, AIC, SC and HQ criteria.

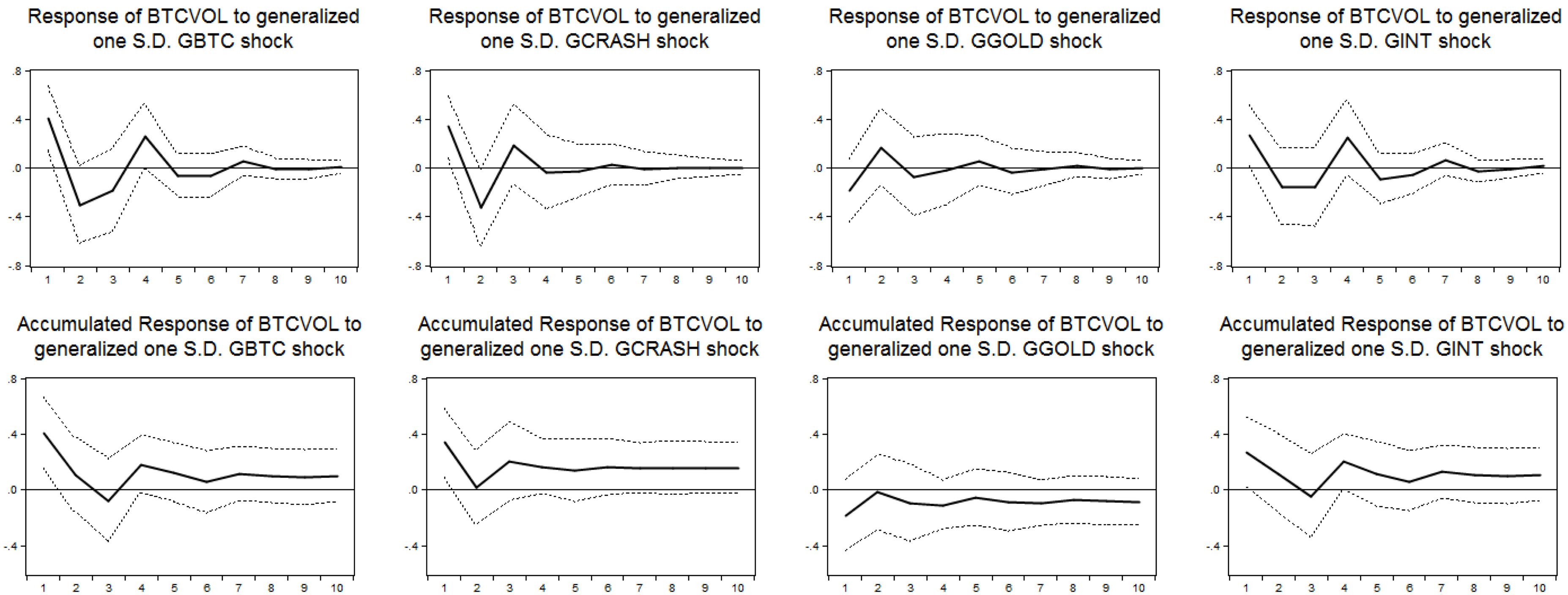

Figure 5 Impulse response and accumulated impulse response functions of bitcoin volatility for one standard deviation shocks to the Google variables BTC, Crash, Crisis, Interest rate, and the S&P and USD currency index volatilities based on data from June 2012 to June 2017. The x-axis represents the time period, in this case 1 equals one week, the y-axis corresponds to the unit of measurement of the bitcoin volatility (log first difference). The impulse corresponds to a one standard deviation shock to the residuals of the variable that is being hit by a shock. The line corresponds to the response, and the dotted lines indicate a 95 % confidence interval of the response. Impulse responses are derived from a VAR(2) model with seven variables; the bitcoin volatility, the Google variables BTC, Crash, Crisis, Interest rate, and the S&P500 volatility and USD currency index volatility. Lag length choices are based on LR, FPE, AIC, SC and HQ criteria.

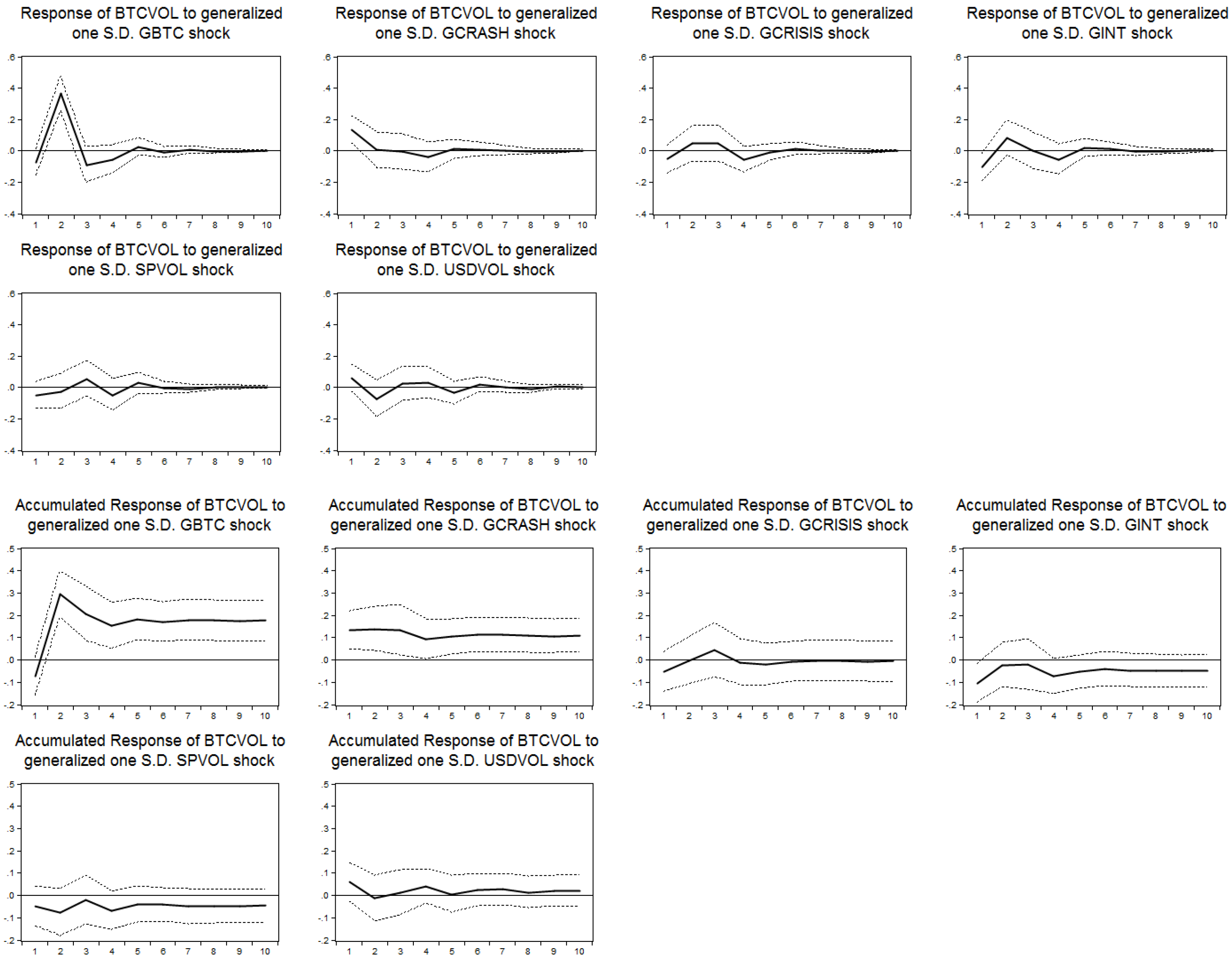

Figure 6 Impulse response and accumulated impulse response functions of bitcoin volatility for one standard deviation shocks to the Google variable BTC and Cyber, and USD currency index volatility, based on data from August 2011 to June 2017. The x-axis represents the time period, in this case 1 equals one month, the y-axis corresponds to the unit of measurement of the bitcoin volatility (log first difference). The impulse corresponds to a one standard deviation shock to the residuals of the variable that is being hit by a shock. The line corresponds to the response, and the dotted lines indicate a 95 % confidence interval of the response. Impulse responses are derived from a VAR(1) model with four variables; the bitcoin volatility, the Google variable BTC and Cyber, and USD currency index volatility. Lag length choices are based on LR, FPE, AIC, SC and HQ criteria.