CESifo WORKING PAPERS

7675
2019

May 2019

Bitcoin and web search query dynamics: is the price driving the hype or is the hype driving the price?

Bernd Süssmuth

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7675*
Category 7: Monetary Policy and International Finance

Bitcoin and web search query dynamics: is the price driving the hype or is the hype driving the price?

Abstract

Using a battery of timely multivariate time series techniques I study the Bitcoin cryptocurrency price series and web search queries with regard to their mutual predictability, Granger-causality and cause-effect delay structure. The Bitcoin is at first treated as a general currency, then as a generic asset. Google queries, although cointegrated, are found to be not helpful in predicting the USD exchange rate of Bitcoin as the speculative bubble in the latter antedates explosive behavior in the former. Chinese Baidu engine queries and compounded Baidu-Google queries predict Bitcoin price dynamics at relatively high frequencies ranging from two to five months. In the other direction, causality runs from the cryptocurrency price to queries statistics across nearly all frequencies. In both directions, the reaction time computed from a phase delay measure for the relevant frequency bands with significant causality ranges from slightly more than one month to about four months.

JEL-Codes: C320, E320, E420, G120, G150.

Keywords: bitcoin, bubbles, frequency domain, causality.

Bernd Süssmuth
Institute for Empirical Research in Economics
University of Leipzig
Grimmaische Str. 12
Germany – 04109 Leipzig
suessmuth@wifa.uni-leipzig.de

Foremost, I would like to thank Jörg Breitung and Helmut Lütkepohl for a host of very valuable comments and suggestions as well as Benjamin Loeper and Jingjing Lyu for their excellent research assistance. My thanks are also due to Felix Noth, Julian Hackinger, and participants of the TUM Research Seminar in Economics for many helpful discussions. The ususal disclaimer applies.

1 Introduction

Terms like cryptocurrency, crypto-token, and blockchain technology are on everyone's lips. The most prominently this applies to Bitcoin (BTC); see Cheah and Fry (2015) and Bariviera et al. (2017). But is this high-pitched online as well as offline media interest already defining what is usually referred to as a hype? According to the *Cambridge Dictionary* the noun hype defines advertising, news reports, and public praise for a new product or service, which is used to make people excited about buying or trying it. Similarly, the *Oxford Dictionary* describes the informal meaning of hype as extravagant or intensive publicity or promotion of something. Meeting these definitions initial crypto-coin offerings (ICO) are frequently vouched for by celebrities such as Paris Hilton relating this form of promotion to dissipative advertising. Dissipative advertising has two crucial characteristics (Linnemer, 2012). First, it contains –if at all– only few informative content. Secondly, it is observable that a substantial amount of money has been spent for it. In this case, the observable expenditure is a celebrity (Paris Hilton) endorsing the cryptocurrency rather than, for instance, an anonymous, less extravagant financial analyst. Some celebrities, like Colombian sports star James Rodriguez, give their name to launch personalized cryptocurrencies in an effort to intensify publicity and make people excited about investing in it. This case (of the "JR10 token") does not represent a form of dissipative advertising. It is rather a kind of brand value promotion as investors enjoy exclusive benefits such as fan-interaction with the star and the possibility of purchasing rare collectionables.

The two sample cases highlight the issue of causal directionality or predictability, which is at the heart of my research agenda, quite well. If the star is dissipatively advertising (Hilton case), the hype is driving the price of the cryptocurrency. However, if the cryptocurrency is used as a merchandising instrument (Rodriguez case), the star freerides on the media attention to currently high-priced cryptocurrencies like the BTC. The price is driving the hype in this case.

Thus, the crucial question to be addressed is: Is the allocation of funds to virtual currencies or risky assets such as the BTC increasing with public interest, attention, and popularity of the financial vehicle or its publicity in general? Or is it the other way around?

Notably, for the verb 'to hype' the *Cambridge Dictionary* provides the following definition. To hype is to make something more exciting or important than it is. Transferring this definition to the context of financial time series, it seems to bear the possibility of speculative bubbles (Homm and Breitung, 2012). A prominent narrative in this context is the case of Nasdaq-listed Long Island Iced Tea Corp (LTEA) in the last quarters of 2017. In the third quarter of 2017 LTEA faced a net loss amounting to USD 3.9 million. When LTEA announced its realignment of business and its re-naming into Long Blockchain Corp (LBCC) on 21 December 2017, its shares soared and tripled in value within hours; see Figure 1. To the present LBCC continues a non-alcoholic beverage subsidiary.

This raises another research question of the present study. If there is evidence for rational speculative bubbles in cryptocurrency series like the BTC exchange rate, can any cointegrated measure of the BTC-hype act as an early warning device for such bubbles?

The related literature published in economics journals, though growing, is still of rather handy size. It comprises, among others, the survey by Böhme et al. (2015), the studies by Cheah and Fry (2015) and Urquhart (2017) treating BTC as standard financial asset prone to speculative bubbles and price clustering, and Brandvold et al. (2015) treating BTC as globally exchange-traded currency. Seminal interdisciplinary contributions emanating either from the information systems or the econophysics field of study include Garcia et al. (2014), Bariviera et al. (2015), Kristoufek (2015), Li and Wang (2017), and Alvarez-Ramirez et al. (2018). The pioneering study on web search queries and BTC prices by Kristoufek (2013) referred to BTC as a currency but treated it rather as an asset. The study has two crucial drawbacks that the present analysis overcomes. First, it only covers the seed phase or early trading period of the BTC from mid-2011 to mid-2013.[1] Secondly, it relies on web search queries on Google and Wikipedia only, which both were and still are blocked in the People's Republic of China and thus inaccessible for a substantial share of investors, users, and miners of BTC at the time and to the present. During November and

[1] An exemplary more recent study, covering the period from January 2013 to February 2018, in the tradition of Kristoufek (2013) is Kjærland et al. (2018). However, it representatively also suffers from relying exclusively on web queries performed by means of the Google web search engine.

December 2013, for example, roughly half of all BTC trades were made in Chinese yuan (Brandvold et al., 2015, p. 20). Ciaian et al. (2018, p. 178) note regarding the regional distribution and the trading currency composition for the BTC that –while the USA and the USD dominated the BTC market in the first years after its introduction– nowadays "almost all [BTC] trading is done in China." The authors document the "staggering rise of China as the dominant trader" of BTC by showing that from less than a ten percent share in January 2012, the yuan made up nearly 100 percent of all BTC trading by the end of 2016. Although this share declined at the latest since fall 2017, when the Chinese government announced to block the access to foreign ICO and crypto-to-fiat exchanges, it seems fair to state that over the period of analysis of the present study (from mid-2011 to the first quarter of 2018) the average share is at least 50 percent. This fact renders Google and Wikipedia series a seriously incomplete measure of attention allocation or (potential) investors' interest.

The present study opts for an integrated approach by taking both perspectives, i.e. BTC as a currency and BTC as an asset, into account. In terms of hype measures, I do not only consider Google trends statistics (provided in normalized terms within my frequency of choice of one month by Google Inc./LLC) for "Bitcoin" searches as, e.g., in Cheah and Fry (2015). I also use the Chinese web search engine Baidu non-normalized query statistics for "比特币" (i.e. "Bitcoin"). The Baidu Zhishu or Baidu Search Index (BSI) is the query statistics of the by far most commonly used web search engine in China given the Google ban that preceded the BTC launch by a couple of months. It reports absolute query figures in monthly frequency. Its informative content has been recently approved by Liu et al. (2016) and He et al. (2018) using it to successfully predict dengue fever outbreaks and HIV incidences in contemporary China, respectively. According to NetMarketShare (`netmarketshare.com`), tracking usage shares of web technologies, the market shares of Google and Baidu in August 2018 amounted to about 70 and 20 percent for mobile devices and 76 and 11 percent for desktop/laptop devices, respectively. None of the other engines even just nears a double digit usage share.

To sum up the crucial point made in the last paragraphs, the previous literature suffers from a substantial deficiency by more or less completely disregarding the regional origin and distribution of BTC-related activity. Ignoring that since mid-2011 about half of BTC trades and investments on average emanate from China lets indicators for attention allocation and investors' curiosity, that are exclusively based on online services such as Google, Wikipedia or Twitter known as blocked or censored in China, appear as unsatisfactory. This concerns all of the most recent and of the most related studies (Kristoufek, 2013, Garcia et al., 2014, Ciaian et al., 2016, Kjærland et al., 2018, Aalborg et al., 2019).

Figure 1: Nasdaq USD opening value of LTEA (LBCC) stock, 2017

It remains to state that the overarching research question is to quantitatively assess how helpful internet data from "secondary sources" (Edelman, 2012) are in predicting and, ultimately, in modeling fintech-related price dynamics at the end of the 2010s.

The methods of choice comprise a battery of multivariate time series techniques in the time and frequency domain such as Chow-type breakpoint testing for speculative bubbles (Phillips et al., 2011; Homm and Breitung, 2012) and testing for Granger causality in the frequency domain (Breitung and Candelon, 2006). They have been successfully applied in diverse contexts; see, e.g., Homm and Breitung (2012), Gómez-González et al. (2014), and Tastan (2015). To assess the corresponding cause-effect reaction time for relevant frequency bands with significant Granger causality I compute a phase delay measure recently developed by Breitung and Schreiber (2018).

2 Treating Bitcoin as a currency

Against the backdrop of the fact hat the number of units of BTC as a digital currency that can be created or "mined" is finite (i.e. is in limited supply) and the entire BTC functioning as decentralized computer network, there is a continuing heated debate of whether there is a fundamental value of this cryptocurrency (e.g. Cheah and Fry, 2015, Yermack, 2015, Andolfatto and Spewak, 2019, Hayes, 2019). While, for instance, Cheah and Fry (2015) find the fundamental price of BTC to equal zero, Hayes (2019) takes a production cost perspective in arguing that the concerted computational effort in BTC mining globally uses electrical power incurring a real monetary cost for producers and, thus, implies a non-zero value of embodied costs of production (on the margin). A theoretical rationalization of this argument results in a non-zero intrinsic value. Hayes (2019) back-tests his pricing model value against the BTC market price and confirms its predictive power for the latter over a period of approximately five years. Andolfatto and Spewak (2019) take a quantity theory of money stance by arguing that the future USD price of BTC will not only depend on the capped supply of BTC, which is in its "orthodox version" (heterodox versions are referred to as "hard forks") capped at 21 million units, but will also crucially depend on the "ever-expanding supply" of alternative cryptocurrencies (altcoins) emanating from the open-source nature of BTC. In their view a "fundamental demand" of BTC as a storage and transfer system provides a non-zero lower bound on the BTC price. It is rooted in permissionless access, decentralized database management, network effects, and the first mover advantage vis-à-vis altcoins.

Whether one judges these positions as rejecting allegations of a zero fundamental value of the BTC or not –in particular, given its finite nature and let alone any normative aspects–[2] is irrelevant for the following argumentation. For BTC being traded above its fundamental value a non-zero fundamental value is simply not a necessary condition (Andolfatto and Spewak, 2019).

[2]These concern not only regulatory and shadow economy issues but possibly also ecological aspects. Recently, for example, Mora et al. (2018) quantify the social cost of BTC mining and usage, under the assumption of a similar rate of adoption of other broadly adopted technologies, to produce enough CO_2 emissions to push global warming above 2 (1.8) degree Celsius (Fahrenheit) within less than three decades.

Treating BTC as a currency, it is straightforward to assess whether its exchange rate dynamics is prone to temporary explosive behavior or speculative, though not necessarily irrational or exuberant, bubble dynamics. The number of studies that do so using historical BTC data is vast. The overall finding is all but diverse. It documents a clear-cut confirmation of (temporary) explosive dynamics or bubbles. The latter have been explored and recently also dated quite frequently in a young but strongly growing strand of literature mostly relying on bubble detection testing; see Cheah and Fry (2015), Cheung et al. (2015), Su et al. (2018), Li et al. (2019), and Hafner (2019). But is it a satisfactory endeavor to ex post date possibly short-lived bubble dynamics? Seen against the backdrop of not addressing the recent blame on economists to not foresee major market crashes or crises, I would suggest to negate and to take up the agenda with a different twist.

Suppose that growth rates of historical BTC exchange rate series, expressing the value of the BTC in US dollars (USD) or in yuan units of Chinese Renminbi (CNY), vary with growth rates of corresponding web search engine queries ("Bitcoin" in case of Google trends and "比特币" in case of the Baidu Zhishu) over time with some feedback, such that they can be tested to be cointegrated over longer periods of time.[3] If then for the queries series an adequate bubble detection test finds a date preceding the date of the one of the exchange rate series, the impending web searches' bubble burst foreshadows the one in the respective exchange rate series. Put it differently, if dynamics in BTC exchange rates and web search queries share a long run stochastic trend and the hype is really driving the price across high and medium term frequencies, short-term forecasts of the dynamics in web search query rates, i.e. of the hype, may act as an early warning device for a bubble burst in the BTC rates. Given the mere popularity of the latter, which sees many hundreds of millions of USD (or CNY) worth of transactions across its system on a daily basis, let the above sketched strategy look more promising than a strategy solely resting on ex post bubble detection testing.

[3]The idea of feedbacks with digital traces of collective social behavior is, generally, also taken up in quantitative studies emanating from computational linguistics and social and information networks; see, e.g., Loughran and MacDonald (2011) and Garcia et al. (2014).

2.1 Data

BTC price series of daily frequency are obtained from CoinDesk (`https://coindesk.com`) denoted in USD. CoinDesk provides the data as Greenwich Mean Time (GMT) end-of-day closing index price, where the latter is aimed to capture the standard retail price reference for industry participants and accounting professionals. It represents an average of leading global BTC exchanges that conform to certain minimum criteria for price discovery and validity.[4] The historical index data commence on July 1, 2013. Any data prior to that date are based on the Mt. Gox price data (see, e.g., Cheung et al., 2015). In order to avoid critical issues of possibly heteroskedastic inter- and intra-weekly regularities in BTC trading –among others, including bias due to weekends, secular holidays, and festivities– my choice of observation frequency is monthly. The conditional heteroskedasticity issue against the backdrop of temporal aggregation will be discussed in somewhat more detail at the end of Section 3.1. Google Trends Statistics (GTS) for search string "Bitcoin"[5] can be retrieved in monthly frequency and solely with the maximum value of monthly reported queries automatically normalized to 100 within the sample period. Thus, I take monthly arithmetic averages of the CoinDesk BTC prices and normalize the resulting series analogously to the GTS series in order to conform prices to normalized query statistics and to avoid to induce spurious cointegration through nonconformity.

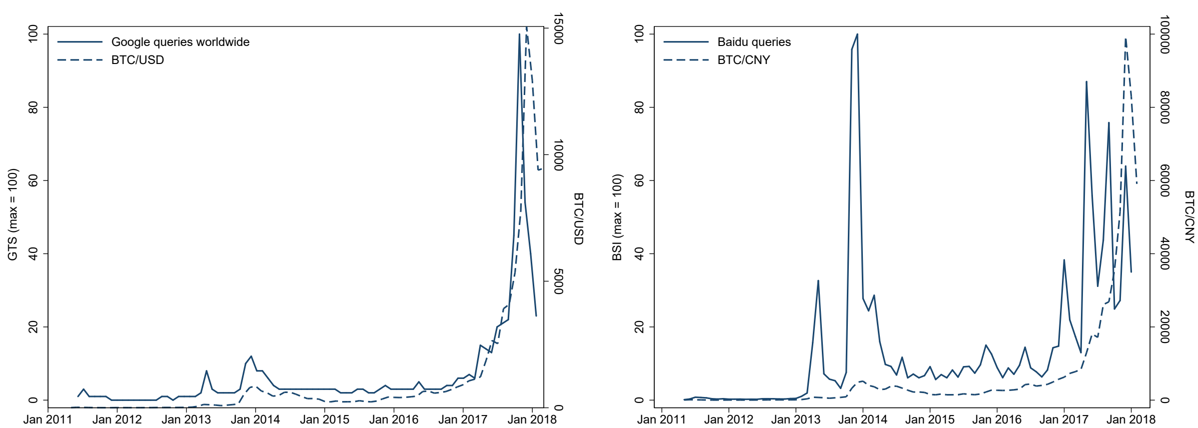
To express the CoinDesk BTC monthly price averages in CNY, i.e. to capture the BTC/CNY exchange rate in quantity quotation, I obtain CNY/USD exchange rate series at monthly frequency from the online database of the Federal Reserve Bank of St. Louis (FRED) coded as EXCHUS. Based on the argument of market liquidity, i.e. "there was practically no liquidity" in the BTC market prior to May 2011 according to the quantitative assessment in Kristoufek (2013, p. 2), the start of the overall sample period is set to May 2011 for both the BTC series and the Baidu Zhishu/Index queries for search string "比特币" (i.e. the Chinese word for"Bitcoin"). The latter is retrieved from

[4]See `https://www.coindesk.com/coindesk-launches-proprietary-bitcoin-price-index`.

[5]As noted in Kristoufek (2013, p. 3), the reported query frequency is not "case sensitive" in the sense that various search string versions of the word (such as "BitCoin" or "bitcoin") are included.

http://zhishu.baidu.com. For comparability reasons, I proceed in analogy to the USD and GTS series and index both series such that the respective monthly maximum value within the sample period corresponds to 100. The overall sample period runs from May 2011 to either January or March 2018. Ending the sample for the mutual BSI queries BTC/CNY exchange rate relationship in January 2018 is justified as by February 2018 the Chinese government announced to block the access to all foreign cryptocurrency exchanges' websites. As the ban took effect during February 2018, it is reasonable to assume that this structural break mostly affected the web search behavior for "比特币" in China, presumably inducing a confounding downward bias in queries statistics. Figure 2 shows in its left schedule the GTS series depicted on the left ordinate and monthly BTC/USD exchange rates in indirect (quantity) quotation, before normalization, depicted on the right ordinate for the period from May 2011 to March 2018. The right schedule displays the normalized BSI series depicted on the left ordinate and the BTC/CNY exchange rate series, also in quantity quotation and before normalization, depicted on the right vertical axis, respectively. The series in this schedule run from May 2011 to January 2018.

Figure 2: Normalized queries and BTC exchange rates, 05/2011 to 01/2018 (03/2018)



Note: Exchange rates in indirect/quantity quotation are based on monthly means of daily closing prices. Sources: FRED, CoinDesk, Google, Baidu; Summary Statistics: Table A.1 (Appendix)

To the best of my knowledge, no reasonable BTC price index or deflator exists in order to transform nominal BTC exchange rates into real terms. The reason is fairly obvious. As

Foley et al. (2019) find in a recent empirical study identifying darknet marketplaces and combining them with seizures of BTC by law enforcement agencies in their projections, about 25 percent of all BTC users are engaged in criminal business. Additionally, the study finds about 50 percent of all transactions in BTC to be related to illegal activities such as drug trade, illegal pornography, and murder-for-hire. As there exist no legal market and corresponding prices for the latter, the construction of an adequate price index or an other reasonable device to express nominal BTC exchange rates in real terms is infeasible.

2.2 Cointegration analysis

As can be seen from the first four rows of Table 1 below, all series in levels as described in Section 2.1 (where BTC/USD and BTC/CNY denote the respective exchange rate in quantity quotation) can be regarded as $I(1)$ processes. This result holds at all conventional levels of statistical significance apart from the GTS in levels for which the Phillips-Perron (PP) test cannot reject the null of a unit root (UR) at the five and at the one percent level of significance only.

Table 1: Unit root (UR) and stationarity test statistics

	ADF	KPSS	PP
BTC/USD	1.994	0.849***	–0.217
Google Trends Statistics (GTS)	1.365	0.603***	–2.880*
Baidu Search Index (BSI)	–1.588	0.278***	–4.178
BTC/CNY	2.204	0.777***	–0.420
Δ(BTC/USD)/(BTC/USD)	–2.157	0.096	–7.812***
ΔGTS/GTS	–2.784*	0.070	–9.017***
ΔBSI/BSI	–8.268**	0.079	–8.246***
Δ(BTC/CNY)/(BTC/CNY)	–2.144	0.096	–7.739***

Note: ADF – Augmented Dickey-Fuller (UR under null); KPSS – Kwiatkowski-Phillips-Schmidt-Shin (stationarity under null); PP – Phillips-Perron (UR under null); * $p < .10$, ** $p < .05$, *** $p < .01$.

On the other hand, the PP test throughout rejects the null of a UR at any conventional level of significance for the series in growth rate expression; see the last column entries in the last four rows of Table 1. The ADF test rejects the null of a UR for the GTS growth rates and BSI growth rates only at the five (and one) percent and one percent level of significance, respectively. Against this backdrop, all series in growth rates can be regarded and treated as covariance stationary and following $I(0)$ processes at reasonably small levels of significance based on three widely used tests with well-known deficiencies.

Table 2: Testing for cointegration I, ADF and Engle Granger (EG) tests

Series (y; x)	Test	Test statistics	1% C.V.	5% C.V.	10% C.V.
BTC/USD; GTS	ADF	–4.190***	–3.535	–2.904	–2.587
	EG	–4.264***	–4.033	–3.412	–3.097
GTS; BTC/USD	ADF	–5.824***	–3.535	–2.904	–2.587
	EG	–5.889***	–4.033	–3.412	–3.097
BTC/CNY; BSI	ADF	0.164	–3.538	–2.906	–2.588
	EG	0.087	–4.037	–3.414	–3.098
BSI; BTC/CNY	ADF	–4.709***	–3.538	–2.906	–2.588
	EG	–4.822***	–4.037	–3.414	–3.098

Note: Test statistics and critical values (C.V.) are of MacKinnon-type;
* $p < .10$, ** $p < .05$, *** $p < .01$

As can be seen from Table 2 and Table 3, and in contrast to earlier work by Kristoufek (2013, pp. 4–5),[6] I find evidence for a cointegrating relationship between the BTC/USD exchange rate series in quantity quotation and the GTS series at any conventional level of significance. Different to the possibly cointegrating relationship between the BTC/CNY exchange rate series in quantity quotation and the BSI series, results for the one between BTC/USD exchange rates and the GTS queries show more unanimity across the different tests for cointegration reported in Table 2 and Table 3. It is a common drawback of the EG approach that it is sometimes not conclusive in identifying which of two variables in the

[6]Series in Kristoufek (2013) are of weekly frequency and run from May 2011 to June 2013.

Table 3: Testing for cointegration II, Johansen-type tests

Series	CI-Rank	Eigenval	Tr-Stats	5% C.V. (Tr)	SBIC	HQIC	AIC
GTS; BTC/USD	0	–	104.692	15.41	11.738	11.632	11.561
	1	0.656	18.369	3.76	10.835°	10.676°	10.569
	2	0.203	–	–	10.663	10.486	10.367
BSI; BTC/CNY	0	–	65.640	15.41	15.430	15.322	15.250
	1	0.437	20.259	3.76	15.021°	14.860°	14.752
	2	0.226	–	–	14.820	14.641	14.520

Note: Information criteria SBIC – Schwarz; HQIC – Hannan-Quinn; AIC – Akaike; ° proposed rank of cointegration (CI); Tr-Stats – trace statistic; Eigenval – eigenvalue (ordering of series is arbitrary); considered lags = 2; period is 07/2017–03/2018 for GTS, BTC/USD; 07/2017–01/2018 for BSI, BTC/CNY.

bivariate case can be used as regressor and why. It thus rests on asymptotic theory that only as sample size goes to infinity, the EG test[7] and the ADF test on the residuals of the long-run (levels) relationship of two differently specified regressions –regarding endogenous and exogenous series– is equivalent for both orderings. For the BTC/CNY exchange rates and the BSI web searches this is not the case as can be seen from the ultimate and penultimate set of tests in Table 2. For BTC/CNY on the right-hand (left-hand) side there is (no) cointegration found by the two tests. Johansen-type tests do not suffer from this deficiency but might be inconclusive in some applications as well. This is also the case here for the standard trace-based Johansen test and the maximum eigenvalue test; see Table 3. Nevertheless, it seems fair to state that there are clear-cut indications in support of a cointegrating relationship between BTC/USD exchange rates and the GTS queries as confirmed by all tests summarized in Table 2 and the Johansen-type test based on an assortment of information criteria, for which two third (SBIC and HQIC) support a cointegrating relationship (Table 3).

For this reason, I focus in the remaining parts of this section on the latter relationship.

[7]The EG test resembles the ADF test. It is performed by regressing the first difference of the residuals of the long-run (levels) relationship on the lagged level of these residuals without a constant. Under the null is no cointegration.

In a first step, I proceed by estimating single equation error correction models (SEECM) of the following form

$$\begin{aligned} \Delta \text{GTS}_t &= \alpha_1 + \beta_1 \left(\Delta \text{BTC/USD}\right)_t + \pi_1 \widehat{\epsilon}_{1,t-1} + e_{1,t} && (1) \\ \Delta \left(\text{BTC/USD}\right)_t &= \alpha_2 + \beta_2 \text{GTS}_t + \pi_2 \widehat{\epsilon}_{2,t-1} + e_{2,t}, && (2) \end{aligned}$$

where $\widehat{\epsilon}_{i,t-1}$, letting subscript $i = 1, 2$ refer to the respective SEECM equation, denotes the lagged error from the long-run (levels) regression, i.e. from the first stage of the EG procedure, and $e_{i,t}$ represents the usual well-behaved i.i.d. error. Corresponding structural parameter estimates are summarized in Table 4.

Table 4: SEECM structural parameter estimates: BTC/USD, GTS

Equation	α_i	β_i	π_i	adj. R-squ.	F(2, 79)
(1)	0.424 (0.328)	0.677^{***} (0.038)	-0.373^{***} (0.077)	0.800	163.43
(2)	-0.396 (0.372)	1.109^{***} (0.056)	-0.836^{***} (0.102)	0.860	250.15

Note: Standard errors given in parantheses; * $p < .10$, ** $p < .05$, *** $p < .01$

Interpreting these estimates, particularly of SEECM adjustment speed π_i, is straightforward. Applying the simple rule of proportion, a value of -0.37 for π_1 in equation (1) means that the cointegrating relationship between GTS_t and $\left(\text{BTC/USD}\right)_t$ (price-to-hype) restores equilibrium about every $100/37 = 2.7$ months. Analogously, $\pi_2 = -0.84$ implies that a pre-period disequilibrium in the hype-to-price relationship is closed in a shorter period of time, that is, every 1.2 months.

2.3 Bubble detection testing of cointegrated series

As argued earlier, I subscribe to the view of Andolfatto and Spewak (2019) that one can accept that the BTC trades above its fundamental value without claiming that its fundamental value is zero or non-zero. Suppose that the fundamental value of a currency, i.e., its price or exchange rate in quantity quotation, equals the present value of its future effective exchange trade prices. Its non-zero or zero fundamental value then equals the present value of future cash flows. In the intrinsically worthless object case, these would be all zero-valued streams. In the following, I widely adopt the notation and follow the argumentation in Homm and Breitung (2012). Discounting with constant risk-free rate (RFR) R and denoting the monetary equivalent of the use value, that is to some extent comparable to a convenience yield of a commodity, of the BTC over time by U renders the no-arbitrage-condition

$$P_t = \frac{E_t\left[P_{t+1} + U_{t+1}\right]}{1+R}. \tag{3}$$

Similar to dividend streams, U can be seen as determined by factors such as competing altcoins, technological factors, acceptance of BTC as a currency or medium of exchange, and the market-to-book ratio, in the sense of the ratio between the market value of the underlying blockchain technology and its replacement costs (Jiang and Lee, 2007; Yermack, 2017; Andolfatto and Spewak, 2019).

Solving (3) by forward iteration, one obtains the usual fundamental value

$$P_t^f = \sum_{i=0}^{\infty} \frac{E_t\left[U_{t+i}\right]}{(1+R)^i}. \tag{4}$$

For a No-Ponze-Game (NPG), (4) converges if

$$\lim_{k\to\infty} E_t\left[\frac{P_{t+k}}{(1+R)^k}\right] = 0. \tag{5}$$

Introducing a bubble component, adds to the no-arbitrage-condition the term $E_t\left[B_{t+1}\right] = (1+R)\,B_t \Leftrightarrow B_t = \frac{E_t[B_{t+1}]}{1+R}$. If this modified condition is violated, B is said to "burst" (Homm/Breitung 2012); else

$$P_t = P_t^f + B_t. \tag{6}$$

For $P_t^f = 0$ and B_t following a random walk with drift, i.e. $\{B_t\}_{t=1}^{\infty} \sim$ RW with drift-parameter $\mu \in \mathbb{R}$,

$$P_t = \mu + \phi P_{t-1} + u_t \;\; \text{with} \;\; \phi = 1,\, u_t \overset{iid}{\sim} N\left(0, \sigma^2\right). \tag{7}$$

The starting point of the bubble detection test by Homm and Breitung (2012) is the forward-iteration of ADF logics, extending the sample by one observation in each step in combination with running recursive regressions, advanced by Philipps et al. (2011):

$$y_t = \mu_y + \phi y_{t-1} + \sum\nolimits_{j=1}^{J} \delta \Delta y_{t-j} + \epsilon_t \;\; \text{with} \tag{8}$$

H_0: $\phi = 1$ (RW) and H_1: $\phi > 1$ (explosive dynamics), and P_t replaced by the more general y_t. Subtracting y_{t-1} on both sides of the equation and disregarding the polynomial autocorrelation-control term $\sum_{j=1}^{J} \delta \Delta y_{t-j}$ renders the Chow-type breakpoint (BP) test for γ by Homm and Breitung (2012)

$$\Delta y_t \;=\; \gamma \left(y_{t-1} D_{\tau T}\right) + \epsilon_t, \text{ where} \tag{9}$$

$$D_{\tau T} \;=\; \begin{cases} 1 & \text{if } t \text{ lies subsequent to } \;\tau T \\ 0 & \text{else,} \end{cases} \tag{10}$$

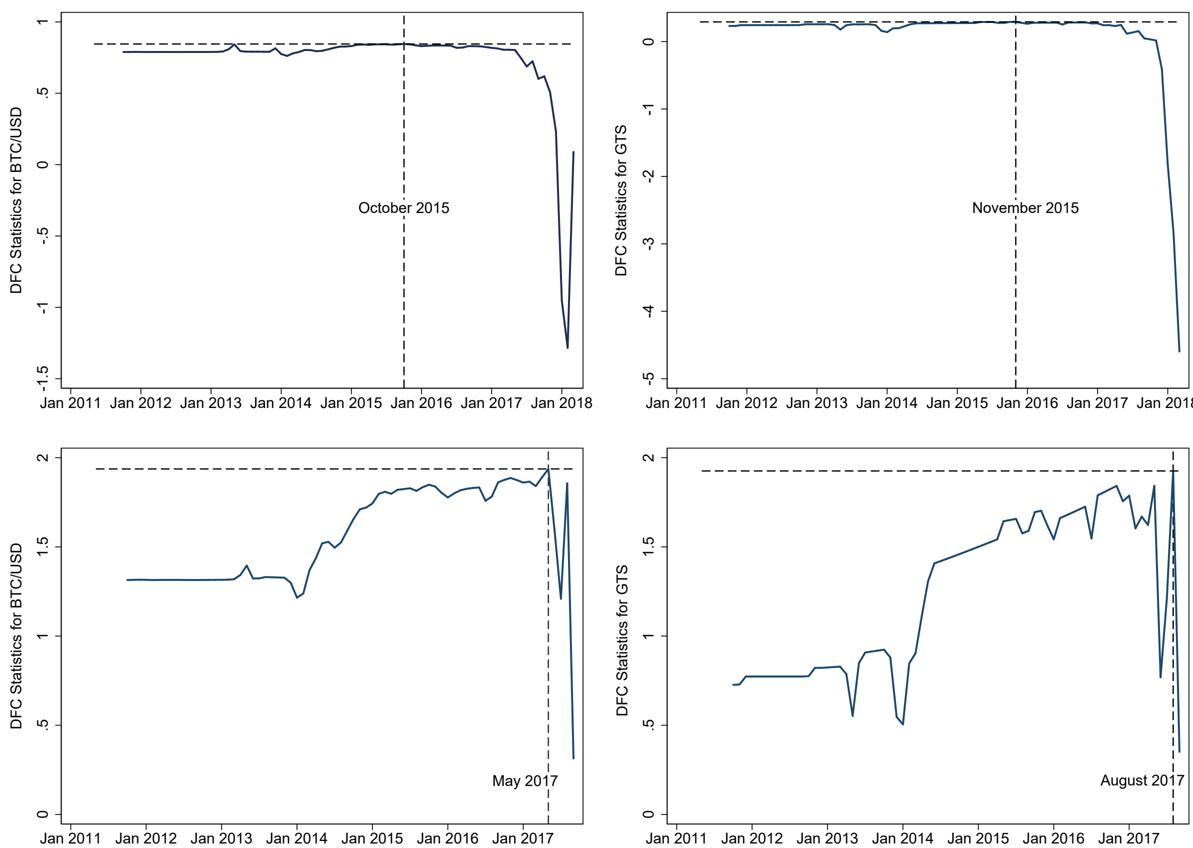
and τT denotes the questioned breakpoint. The BP test statistic for bubble burst identification proposed by Homm and Breitung (2012) is referred to as Dickey-Fuller-Chow (DFC) and reads

$$\underset{\{\tau\}}{\arg\max} DFC_\tau = \frac{\sqrt{T-2} \sum\limits_{t=\tau T+1}^{T} \Delta y_t y_{t-1}}{\sqrt{\sum\limits_{t=2}^{T} \left(\Delta y_t - \widehat{\gamma} y_{t-1} D_{\tau T}\right)^2 \sum\limits_{t=\tau T+1}^{T} \Delta y_t^2}}. \tag{11}$$

Following the strategy proposed by Homm and Breitung (2012), I run the test on the growth rates of BTC/USD and GTS series, respectively. The outcome is graphically summarized in Figure 3.

The vertical dashed lines in the four schedules of Figure 3 identify the suprema of the respective DFC Statistics. The supremum of the DFC Statistics for the GTS series (upper right diagram) does not antedate the DFC Statistics' peak of the cointegrated BTC/USD series (upper left diagram). The occurrence of the BTC/USD bubble in October 2015 is followed by a bubble formation in GTS web search queries in November 2015, that is, with a delay of one month. Thus, the latter does not qualify as an early warning device for the former.

Figure 3: DFC Statistics for bubble detection in growth rates of BTC/USD and GTS



Note: For first (second) row of diagrams sample is 06/2011 to 03/2018 (06/2011 to 09/2017)

As the DFC test starts accumulated at the end of the sample period, beginning it in the immediate aftermath of a crash is suboptimal. One may be concerned that this circumstance also plagues the bubble detection based on DFC Statistics shown in the first row of diagrams in Figure 3. Thus, I also consider a narrowed period running from June 2011 to September 2017 that clearly excludes the crash (Figure 2). Results are shown in the second row of diagrams in Figure 3. They confirm the general finding of the supremum of the DFC Statistics for the BTC/USD growth rates (lower left diagram) preceding the corresponding one of the cointegrated GTS growth rates (lower right diagram). However, according to this DFC test run with narrowed observation period, the BTC/USD bubble is dated in May 2017 and followed by a bubble formation in the GTS rates in August 2017, that is, with a delay of one quarter. The latter span of one quarter is also confirmed by the recursive ADF Statistics of Philipps et al. (2011). It identifies a break date in the BTC/USD growth rates in February 2017 at a five and at a ten percent level of significance. The corresponding result for GTS growth rates is not as clear-cut as two dates, May 2017 and November 2017, turn out significant though at a ten percent level only.

3 Treating Bitcoin as an asset

Recently, there are several indications speaking in favor of treating the BTC rather as asset, in the sense of an actual investment or a precursor of an investment, i.e. neither in the narrow sense of a security nor of a commodity, than as a (digital) currency. These include the discussion of BTC as part of the underlying technology of (near) future investment opportunities such as the so-called 'internet of payments' for the 'internet of things' as, for example, in future versions of Amazon's "dash button" or any form of the blockchain economy in general. An indication pointing in this direction is the registering of domain names `amazonbitcoin.com` in March 2013 and `amazoncryptocurrency.com` in October 2017 by Amazon Technologies Inc.; see, e.g., `https://www.whois.com/whois/amazonbitcoin.com`. Additionally, there exist a growing number of financial market securities that are based on the BTC as central underlyer such as Australian Apollo Capital Fund launched in February 2018, the Postera fund launched in Liechtenstein in March 2018, or the different products and services provided by Crypto Fund AG approved in 2018 by Switzerland's principal stock exchange (Six Swiss Exchange) and the Swiss Financial Market Supervisory Authority (FINMA). For the manifold future role of BTC and its current role as a pioneering new technology in the financial industry in general see Yermack (2017).

Since the dynamics in the focus of this section hinge, as opposed to the preceding section, not on a long run equilibrium relationship but rather represent dynamic relationships at business cycle, or even higher, frequencies, cointegration is not an issue.

Thus, a standard vector autoregressive (VAR) process of order p, i.e. a VAR(p), generating two series x_t and y_t, in reduced form will serve as methodological starting point. That is

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = Dd_t + \sum_{i=1}^{p} \begin{pmatrix} a_{11,i} & a_{12,i} \\ a_{21,i} & a_{22,i} \end{pmatrix} \begin{pmatrix} x_{t-i} \\ y_{t-i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{xt} \\ \varepsilon_{yt} \end{pmatrix}, \tag{12}$$

where d_t denotes a deterministic vector containing constants, trend component, and possibly other exogenous variables,[8] $a_{kl,i}$ denote coefficients and $[\varepsilon_{xt}, \varepsilon_{yt}]'$ serially uncorrelated reduced form errors.

In this bivariate system, Granger-causality or predictability is defined as follows. There is no Granger-causality given if the prediction of x is not improved by lagged values of y. As y does not help to predict x, it is not "granger-causing" y. Thus, if $a_{12,i} = 0$ for $i = 1, ..., p$, y is found to be not Granger-causal for x. Therefore, the adequate choice of tests stems from the F/χ^2 class and, as applied in block-form, represents Wald tests. In general, cointegration implies Granger-causality but not the other way around.

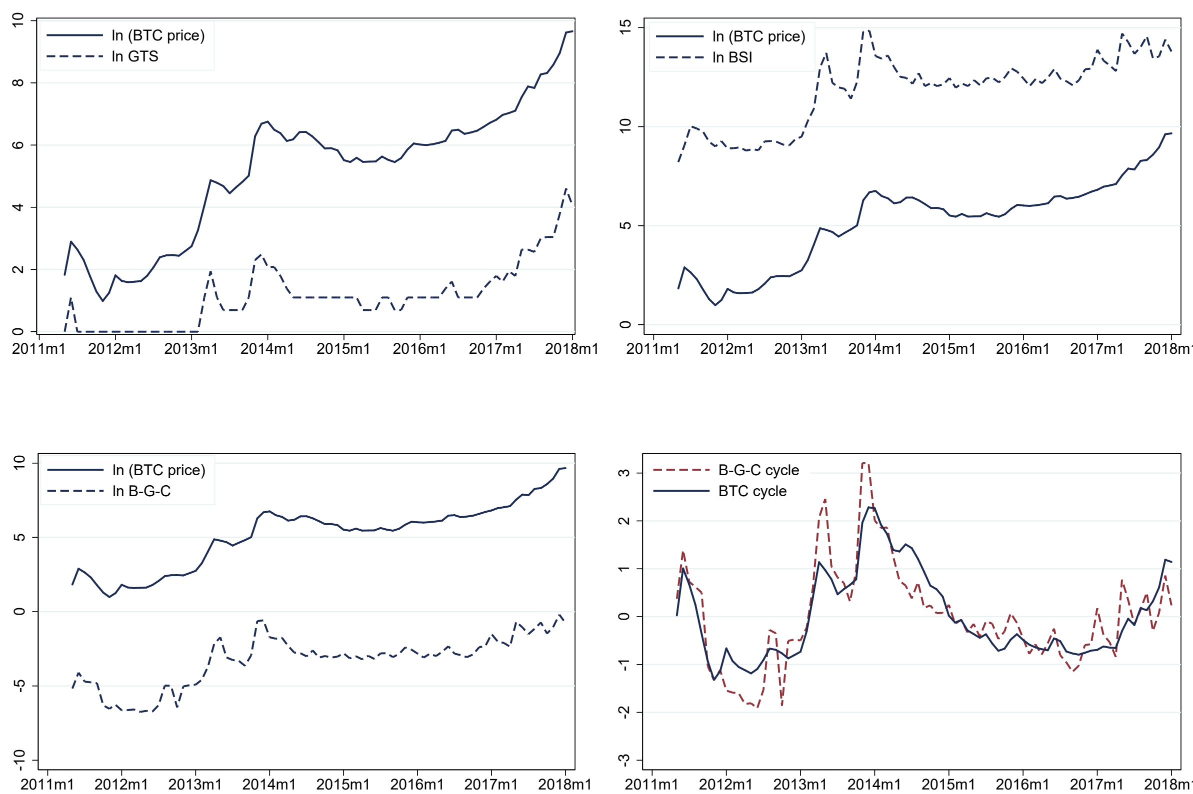
3.1 Data

In contrast to the preceding section treating BTC as a currency, there is no longer a necessity to separate Google from Baidu queries (i.e. GTS from BSI series) as alternative x_t series in a bivariate VAR specification as detailed in the preceding subsection. This

[8] The recent findings of Kjærland et al. (2018) speak in favor of a parsimonious specification of these exogenous factors as these authors identify no significant impact neither from the technological factor "hashrate" (i.e. the variable speed at which a computer can complete an operation in the BTC code) nor from oil or gold. Additionally and in contrast to autoregressive terms and publicity measured by –in the present study also allowed to be acting as endogenous– Google web search queries, a popular volatility index (VIX) and the BTC transaction volume turn out throughout their specifications as statistically insignificant exogenous determinants of BTC return dynamics.

is due to the fact that BTC price series need not to be analyzed as exchange rate series in either USD or in yuan units of CNY any longer. It also seems straightforward to compound the two query series to a composite index. Even if this strategy would come at the cost of a dilution of the information contained in the raw BSI series, a Baidu-Google composite (B-G-C) index of queries simplifies the analysis.

Figure 4: Constituent and composite queries benchmarked to BTC price, 05/2011–01/2018



Note: Log levels of BTC price, GTS-, BSI-, B-G-C-queries; BTC price, B-G-C-queries cycle components

Sources: CoinDesk, Google, Baidu; Summary Statistics: Table A.1 (Appendix)

As argued above, a B-G-C index captures about 90 percent of total web search engine usage in recent times. Due to a potential structural break in the series induced by Chinese governmental blocking of the access to non-domestic online BTC trading platforms in February 2018, the period of analysis is in the following restricted to run from May 2011 to January 2018. An equal weightening in the construction of the B-G-C index is justified as in the period of observation, on average, about half of BTC trades and investments emanate from China. BSI queries are normalized in the same way as are the retrieved GTS series summed up and divided by 200; see the upper row of diagrams and the lower left diagram in Figure 4. The lower right diagram shows the series in the lower left diagram with a linear trend removed.

In Section 2.1 it is argued that a monthly frequency of series is chosen to avoid the issue of conditional heteroskedastic (CH) effects. As shown in Drost and Nijman (1993) low order GARCH processes are not closed in the sense of surviving an increasing sampling interval (i.e. the temporal aggregation of underlying time series). If price and popularity series result from aggregating more and more CH effects disappear. Generally, for example, monthly exchange rate series are found to be homoskedastic while corresponding daily and weekly series are not; see Baillie and Bollerslev (1989). Hafner (2008, p. 476) shows that this general result of unclosedness also holds for multivariate, including the present case of bivariate, processes if series represent flow variables. This reasoning lets me refrain from considering CH effects in the following VAR estimations.

3.2 Spectral Granger-causality tests

In the following, I consider Breitung-Candelon-Granger- (BCG-) causality testing in the frequency domain (Breitung and Candelon, 2012) as a more adequate method in the context of the present study. This is due to its potential to continuously quantify mutual predictability at classical business cycle and higher frequencies.

Any n-dimensional stationary process $\mathbf{Y}_t$ has a spectral representation at frequencies $\omega \in [-\pi, \pi]$ in the form of a spectral density matrix $\mathbf{F}(\omega)$.

It is given by the Fourier transform of the covariance function $\gamma_{jk}(\tau)$, $\tau = 0, \pm 1, \pm 2, ...,$ for all $j = 1, ..., n$; $k = 1, ..., n$. As $\mathbf{F}(\omega)$ is even, it is sufficient to examine it in $[0, \pi]$.

It can be written as

$$
\begin{aligned}
\mathbf{F}(\omega) &= \frac{1}{2\pi}\sum_{\tau=-\infty}^{+\infty}\mathbf{G}(\tau)\,e^{-i\omega\tau}, \quad -\pi \leq \omega \leq \pi \quad \text{with} && (13)\\
\mathbf{G}(\omega) &= \begin{pmatrix} \gamma_{11}(\omega) & \cdots & \gamma_{1n}(\omega) \\ \vdots & \ddots & \vdots \\ \gamma_{n1}(\omega) & \cdots & \gamma_{nn}(\omega) \end{pmatrix} \text{and } \mathbf{F}(\omega) = \begin{pmatrix} f_{11}(\omega) & \cdots & f_{1n}(\omega) \\ \vdots & \ddots & \vdots \\ f_{n1}(\omega) & \cdots & f_{nn}(\omega) \end{pmatrix}.
\end{aligned}
$$

An implementation of (13) can be achieved by a VAR estimation with coefficient matrix $\mathbf{A}$. In this case,

$$
\mathbf{F}(\omega) = \frac{1}{2\pi}\mathbf{A}(\omega)^{-1}\boldsymbol{\Sigma}\mathbf{A}(\omega)^{-*}, \text{ where} \tag{14}
$$

$\boldsymbol{\Sigma}$ is the positive-definite covariance matrix of errors; $\mathbf{A}(\omega)$ denotes the Fourier-transform of matrix lag polynomial $\mathbf{A}(L)$ and '*' its conjugate complex transpose, respectively. According to Wold's theorem

$$
\boldsymbol{\Gamma}(L)\,\mathbf{Y}_t = \varepsilon_t \Leftrightarrow \begin{pmatrix} x_t \\ y_t \end{pmatrix} = \Psi(L)\begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}, \text{ where} \tag{15}
$$

η_{jt} denotes Choleski factorized errors and $\Psi(L) = \widetilde{\boldsymbol{\Gamma}}^{-1}$. Geweke (1982) is the first to make use of this property by equating a respective measure of linear feedback to zero under the null in a test of linear dependence and feedback between multiple time series. That is, for instance, for testing for linear feedback to run from y to x across all $\omega \in [0, \pi]$

$$
H_0\colon \quad M_{y\to x}(\omega) = \log\left(1 + \frac{|\psi_{12}(e^{-i\omega})|^2}{|\psi_{11}(e^{-i\omega})|^2}\right) = 0, \text{ where} \tag{16}
$$

$y \to x$ denotes y is helpful in predicting x.

Analogously, one may rewrite (12) using lag polynomial notation as

$$\begin{bmatrix} a_{11}(L) & a_{12}(L) \\ a_{21}(L) & a_{22}(L) \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \varepsilon_{xt} \\ \varepsilon_{yt} \end{bmatrix}. \tag{17}$$

In this notation, x_t is not Granger-causal for y_t if H_0: $a_{21}(L) = 0$. To test this hypothesis, the second equation of system (17), that is,

$$\begin{aligned} y_t &= \alpha_1 y_{t-1} + ... + \alpha_p y_{t-p} + \beta_1 x_{t-1} + ... + \beta_p x_{t-p} + \varepsilon_{yt} \\ &= \alpha(L) y_{t-1} + \beta(L) x_{t-1} + \varepsilon_{yt}, \end{aligned} \tag{18}$$

where $\alpha_i = a_{22,i}$ and $\beta_i = -a_{21,i}$, is used. Following Breitung and Candelon (2006), the definition of the falsified BCG-causality under the null then reads as follows.

Definition 1 *x_t is not a cause of y_t at frequency ω if the gain function of the filter $\beta(L)$ equals zero at frequency ω, that is,*

$$H_0\text{:} \ \left|\beta e^{\omega i}\right| = \left|\sum_{j=1}^{p} \beta_j \cos(j\omega) + \sum_{j=1}^{p} \beta_j \sin(j\omega)\, i\right| = 0.$$

The necessary and sufficient conditions for the inexistence of BCG-causality running from x_t to y_t, thus, are $\sum_{j=1}^{p} \beta_j \cos(j\omega) = 0 \wedge \sum_{j=1}^{p} \beta_j \sin(j\omega) = 0$. The usual F-test logic for a linear combination of $R(\omega)\beta = 0$ applies.

For the standard representation of a VAR(p) extended to a VARX(p), as shown in (12) above, it is straightforward to extend the framework of spectral Granger causality to the case of an additional exogenous variable z, that is, typically to the case of a deterministic trend function

$$x_t = c_1 + \sum_{j=1}^{p} \alpha_j x_{t-j} + \sum_{j=1}^{p} \beta_j y_{t-j} + \sum_{j=1}^{p} \delta_j z_{t-j} + e_t. \tag{19}$$

It turns null hypothesis (16) into a conditional one, that is, into H_0: $M_{y \to x|z}(\omega) = 0$ (Tastan, 2015, pp. 1159-1160). A mere eyeballing of time series shown in Figure 2 is suggestive for the BTC as well as the queries level series to follow an exponential trend. For natural log expressions of the corresponding levels series (Figure 4), it thus appears appropriate to test for a linear trend to be considered in z. The VAR approach generally presupposes stationarity of included series or the use of first-differenced $I(1)$ variables requiring a comprehensive range of pretests for cointegrating relationships. The modified Wald test for Granger causality proposed by Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996) does not necessitate pretesting for cointegration. These authors suggest that the usual Wald statistics will be valid for Granger causality on p lags of a variable in an overfitted VAR$(p + d_{\max})$ with $d_{\max}$ denoting the highest order of integration in the system of possibly integrated processes. Breitung and Candelon (2006) propose this strategy to be also followed in the present frequency domain context. Assuming that $d_{\max} > 0$, the corresponding test regression is written as

$$x_t = c_1 + \sum_{j=1}^{p} (\alpha_j x_{t-j} + \beta_j y_{t-j}) + \sum_{k=p+1}^{p+d_{\max}} (\alpha_k x_{t-k} + \beta_k y_{t-k}) + e_t, \tag{20}$$

where null hypothesis $M_{y\to x}(\omega) = 0$ involves β_j for $j = 1, ..., p$ only. It can be tested using the standard Wald statistics. As the coefficients on the additional lagged variables are not included in the computation of the test statistics, the same χ^2-distributed Wald statistics can be used as in the case without lag-augmentation. The same applies to the VARX$(p + d_{\max})$ model

$$x_t = c_1 + \sum_{j=1}^{p} (\alpha_j x_{t-j} + \beta_j y_{t-j} + \delta_j z_{t-j}) + \sum_{k=p+1}^{p+d_{\max}} (\alpha_k x_{t-k} + \beta_k y_{t-k}) + e_t, \tag{21}$$

where x and y series represent natural log expressions of levels, z denotes a deterministic linear trend, and H_0: $M_{y\to x|z}(\omega) = 0$, that I will also use in the following to test for spectral Granger-causality.

3.3 Spectral delay assessment

The diagonal elements $f_{11}(\omega), ..., f_{nn}(\omega)$ of $\mathbf{F}(\omega)$ in (13) are the real valued autospectra or power spectra. The off-diagonal elements represent cross spectra $f_{jk}(\omega) = c_{jk}(\omega) - iq_{jk}(\omega)$, consisting of $c_{jk}(\omega)$ cospectra and $q_{jk}(\omega)$ quadrature spectra. The cross-spectra are complex valued functions in ω, but simple manipulations yield the more readily interpretable real phase shift measure

$$\phi_{jk}(\omega) = \arctan \frac{-q_{jk}(\omega)}{c_{jk}(\omega)}. \tag{22}$$

Both, in the autospectral and in the pairwise bivariate case, $\phi(\omega)$ and $\phi_{jk}(\omega) = \phi_{xy}(\omega)$, the phase shift can be visualized either on circular scale or on linear scale or on both scales at a time. For the latter see, e.g., Heer and Süssmuth (2013, p. 406). As the phase shift corresponds to the phase angle –also referred to as angular coordinate or polar angle– in the circular space, it repeats every 2π periods due to the circular diameter equaling 2π. It is, thus, said to be only defined mod 2π. Additionally, as illustrated in detail in Heer and Süssmuth (2013), there is a frequency, where the phase shift reaches π, that is, where the counterclockwisely rotated phase angle coincides with the horizontal originating from the circle center. It is this phase shift that cannot be distinguished from $\phi_{xy}(\omega) = -\pi$. In this case, there is no difference between the statement "series y leads series x with half a cycle length" and the statement "series y lags series x with half a cycle length." A discontinuity at the angular frequency corresponding to this phase shift results.

A solution using the four-quadrant version of the inverse tangent, i.e. of the arctan, function that is usually referred to as atan2 is recently proposed by Breitung and Schreiber (2018, p. 64). Reconsider VAR representation (18) above and rearrange, given invertibility of $\alpha(L) = 1 - \sum_{j=1}^{p} \alpha_j L^j$, to obtain

$$y_t = \frac{\beta(L)}{\alpha(L)} x_{t-1} + \nu_t = \rho(L) x_t + \nu_t, \tag{23}$$

where $\rho(L) = \beta(L) L / \alpha(L)$ and $\nu_t = \alpha(L)^{-1} \varepsilon_{yt}$. According to Breitung and Schreiber (2018) the phase shift induced by filter $\rho(L)$, for non-zero gains of implied filters $\beta(L) L$ and $\alpha(L)$ and for the atan2 definition space $(0, 2\pi]$, is given by

$$\phi_\rho(\omega) = \text{atan2}\left(\frac{q_\rho(\omega)}{c_\rho(\omega)},\ \text{sgn}[q_\rho(\omega)],\ \text{sgn}[c_\rho(\omega)]\right), \tag{24}$$

where, for two general arguments $(a,\ b)$ atan2 in terms of the standard arctan function can be expressed as follows

$$\operatorname{atan2}(a,b)=\begin{cases}\arctan\left(\frac{a}{b}\right) & \text{if } a>0,\\ \arctan\left(\frac{a}{b}\right)+2\pi & \text{if } a<0 \text{ and } b\geq 0,\\ 2\pi & \text{if } a=0 \text{ and } b>0,\\ \text{undefined} & \text{if } a>0 \text{ and } b=0.\end{cases} \tag{25}$$

Using (24) in combination with (25), the spectral delay measure by Breitung and Schreiber (2018) is calculated as

$$d(\omega)=\frac{\widetilde{\phi}_{\rho}(\omega)}{\omega}, \tag{26}$$

where $\widetilde{\phi}_{\rho}(\omega)$ denotes "unwrapped" phase delay. With phase unwrapping, Breitung and Schreiber (2018, pp. 64–65) refer to the following strategy. If at some frequency ω the term $q_{\rho}(\omega)$ switches its sign for $c_{\rho}(\omega)>0$ the resulting phase shift will show a discontinuous jump down from (or up to) 2π to (or from) an arbitrarily close to zero value. As the phase shift in principle is only identified up to adding integer multiples of 2π, the implied delay function will jump between $2\pi/\omega$ and zero; see (25) in combination with (26). By definition, however, all phase shifts measured by (24) are to be mapped into the interval $(0,2\pi]$. The "unwrapping" workaround is to remove discontinuities in the phase shift function by adding or subtracting integer multiples of 2π where needed. As the unwrapping procedure is independent of the estimation of $\phi_{\rho}(\omega)$, it does not bias the sampling uncertainty of locally identified measures.

In analogy to the general requirement of the underlying VAR in the preceding subsection to be of order $p\geq 2$, a VAR order in excess of two lags, i.e. $p\geq 3$, is necessary for estimating spectral delay; see Breitung and Schreiber (2018).

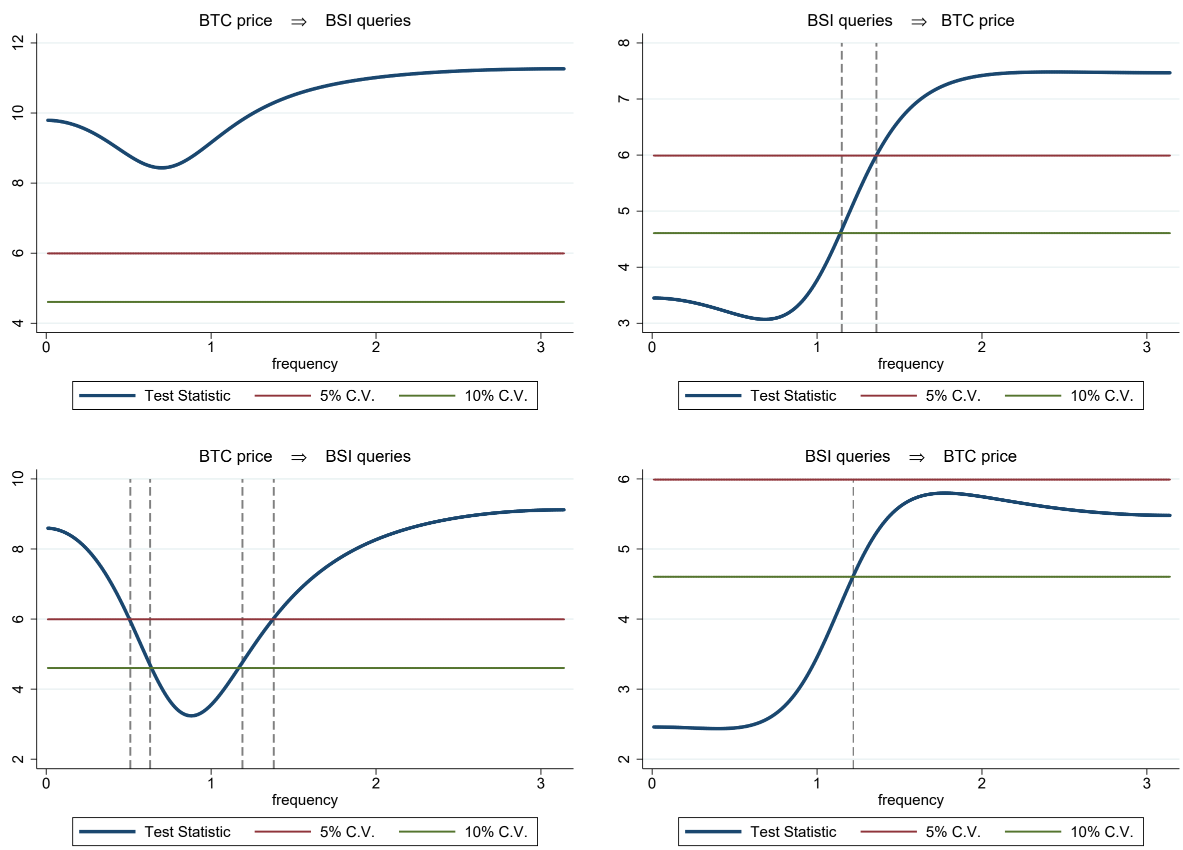
3.4 Findings and interpretation

As argued above, the starting point of the analysis is the estimation of a bivariate VAR(p) system as specified by (12), (15) and (17), (18) considering a linear trend as exogenous variable as captured by Dd_t in (12) or analogously by z in (19). Throughout the natural log transformed BTC price series in USD serves as x_t series. For y_t I first consider natural log expressions of absolute BSI queries (dashed series in the upper left window of Figure 4) and as an alternative ln-transformed B-G-C series (dashed series in the lower left schedule of Figure 4), which by construction are logs of a normalized time series, i.e. logs of values from the $(0, 1)$ interval. All estimated VAR models satisfy the stability condition. Additionally, the lag order of the respective VARX model that minimizes for both considered y_t series the AIC and final prediction error (FPE) information criterion is $p = 3$. A standard F-test finds the linear trend as exogenous variable in the VARX(3) to be significantly different from zero with a p-value of 0.04 for BSI queries and a p-value of 0.05 for the B-G-C composite series, respectively. The outcome of BCG-causality tests (Section 3.2), reported and visualized in Figure 5 and Figure 6 as well as corresponding phase delay measures shown in Figure 7 and Figure 8, rely on estimates of corresponding bivariate VARX(3) models.

To correctly interpret the diagrams in the four figures, it is important to note that frequency depicted on the respective abscissa in the respective four schedules refers to angular frequency. Therefore, the highest measurable frequency is $\omega^{\max} = \pi = 3.1415...$ corresponding to an ordinary frequency of $f^{\max} = \frac{\omega^{\max}}{2\pi} = \frac{\pi}{2\pi} = 0.5$. It is referred to as Nyquist frequency and represents the lowest discernible periodicity of a contained cyclic mode $P^{\min} = (f^{\max})^{-1} = 2$, i.e. a two-period (2 months) cyclicality. Additionally, note that the dashed confidence bands displayed in Figure 7 and Figure 8 are computed for a five percent level of significance.

For standard BCG testing the BTC price is found to be helpful in predicting BSI queries across all frequencies (upper left diagram in Figure 5). Relying on the lag-augmented (Toda-Yamamoto modified) BCG test, shown in the lower left diagram of Figure 5, this is also the case apart from the (0.51, 1.38) frequency band corresponding to periodicities of 4.6 to 12.3 months for the five percent level of significance and from the (0.63, 1.21) frequency band corresponding to periodicities of 5.2 to 10 months for the ten percent level of significance. See the four thresholds given by dashed vertical lines in the lower left diagram identifying the respective bands in Figure 5.

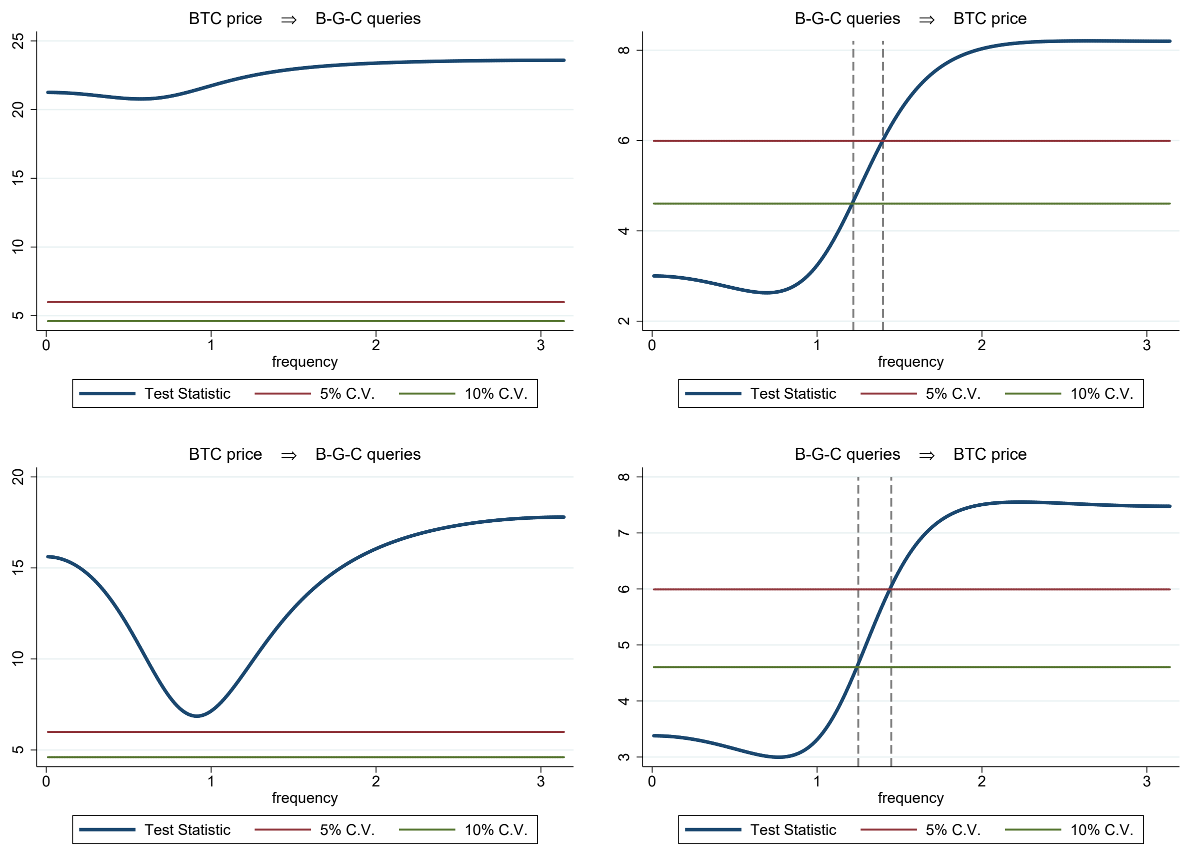
Figure 5: Spectral Granger causality tests using BSI queries as hype measure



Note: First row of diagrams: standard BCG tests; second row: Toda-Yamamoto modified BCG tests; left column of diagrams: price-to-hype (BTC → BSI), right column: hype-to-price (BSI → BTC)

Combining this insight with results on spectral delay reported in the first column of Figure 7 reveals that the corresponding delay between cause (BTC price) and effect (BSI queries) is significantly different from zero for periodicities lower than six and four months (corresponding to abscissa values of about 1 and 1.5, respectively). It is of relatively low magnitude, that is, it amounts to just about one month.

Figure 6: Spectral Granger causality tests using B-G-C queries as hype measure



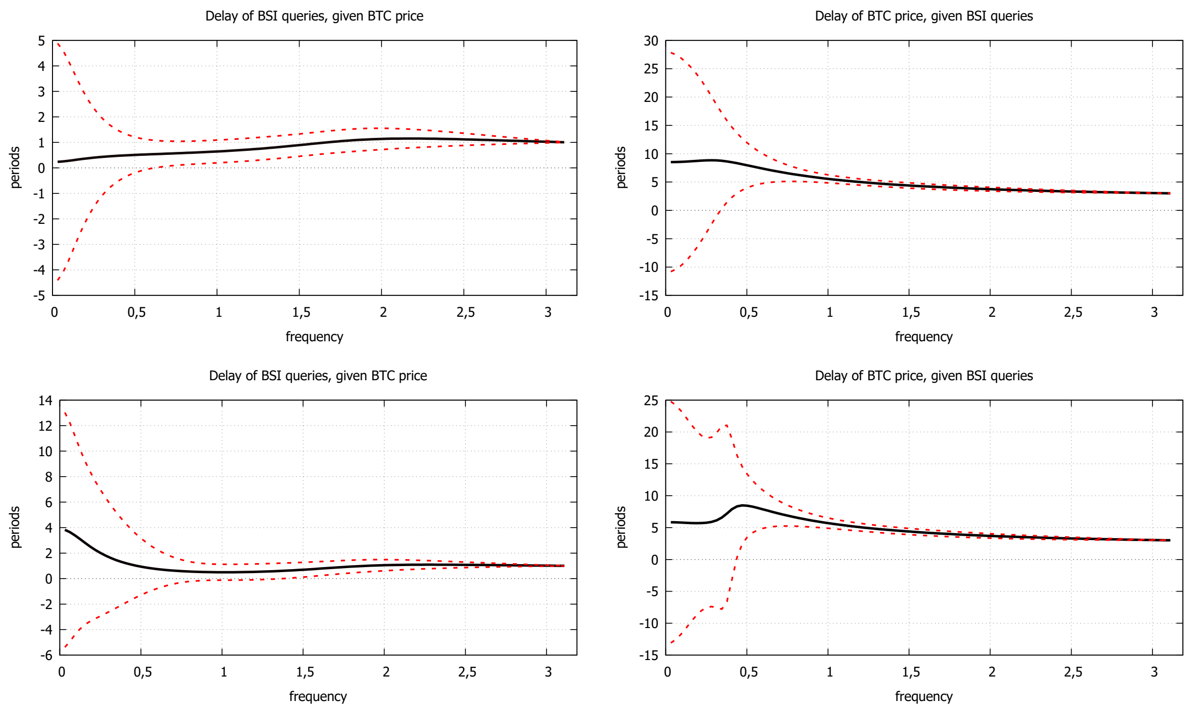
Note: First row of diagrams: standard BCG tests; second row: Toda-Yamamoto modified BCG tests; left column of diagrams: price-to-hype (BTC → B-G-C), right column: hype-to-price (B-G-C → BTC)

Regarding the second column of diagrams in Figure 5 and Figure 7, it can be noted that the hype measure BSI queries is Granger-causal for the BTC price series only for frequencies (periodicities) in excess of 1.15 (below 5.5 months) for the ten percent level of significance[9] and standard BCG testing with a delay of about three to four months.

For the modified BCG test result, there is no Granger-causality found between BSI queries and BTC price that is significant at the five percent level of significance (lower right schedule in Figure 5). For the ten percent level, the threshold resembles the one of the one found by means of standard BCG testing: 1.22 (5.2 months). The same holds for the corresponding phase delay; see the lower right schedule in Figure 7.

[9] At the five percent level of significance the frequency (periodicity) threshold is 1.36 (3.5 months).

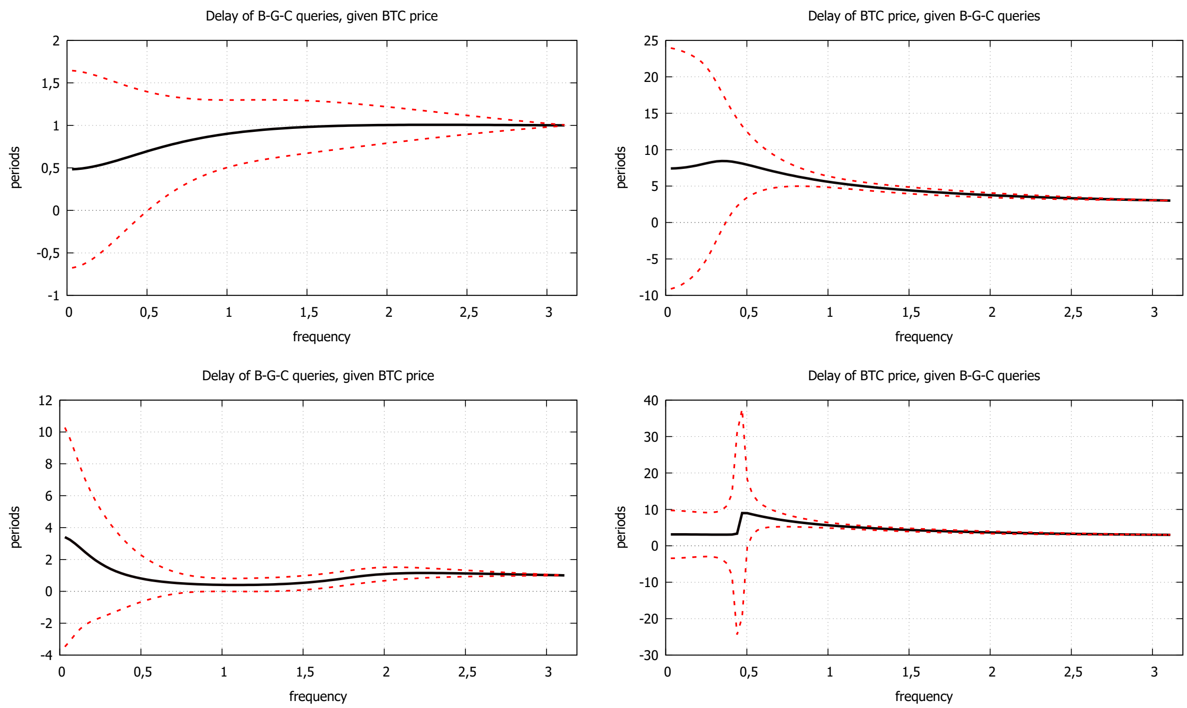
Figure 7: Spectral delay estimates using BSI queries as hype measure



Note: First row of diagrams: based on standard bivariate VAR; second row: lag-augmented VAR based; left column of diagrams: price-to-hype (BTC → BSI), right column: hype-to-price (BSI → BTC)

For the composite B-G-C series as hype measure, results are visualized in Figure 6 and Figure 8. Focusing on the respective first column of diagrams at first, it can be stated that the BTC price granger-causes the B-G-C hype measure for all frequencies/periodicities, even at a conservative five percent level of significance with a delay of ≤ 1 month. The delay is significantly different from zero at a five percent level for cyclic components with period lengths implying a peak-to-peak distance that is not exceeding half a year or, at least, is lower than six months. Turning to the respective second column of diagrams in Figure 6 and Figure 8, the results for the BSI attention measure are generally also confirmed for the B-G-C queries inasmuch as the hype is causing the price significantly only at high frequencies corresponding to fluctuations with periodicities lower than four to five months. Also in accordance with preceding test results, the phase delay at relevant frequencies amounts to about two to four months.

Figure 8: Spectral delay estimates using B-G-C queries as hype measure



Note: First row of diagrams: based on standard bivariate VAR; second row: lag-augmented VAR based; left column of diagrams: price-to-hype (BTC $\rightarrow$ B-G-C), right column: hype-to-price (B-G-C $\rightarrow$ BTC)

To sum up, while the price in general is helpful in predicting ("driving") the hype more or less immediately, i.e. delayed at maximum by about one month, the hype drives the price only for relatively high frequency dynamics and with a two- to four-times higher delay.

4 Conclusion

This study is the first to overcome a substantial deficiency of the previous literature on BTC price dynamics and web search statistics, which by now has more or less completely ignored the regional origin and distribution of BTC-related activity. Since mid-2011 about half of BTC trades and investments on average emanate from China, motivating my use of Chinese Baidu web searches for "比特币" (i.e. "Bitcoin") and an unweighted Baidu-Google queries composite series as a measure of attention or, colloquially, the hype.

Economic theory (e.g. Manuelli and Peck, 1990) predicts that the price dynamics of an unbacked asset is inherently unforecastable. The present study confirms this only in parts as on the one hand Google queries, although cointegrated with monthly BTC price series, are found to be not helpful in predicting the price series as the speculative bubble in the price series antedates explosive behavior in the web search series by up to three months. Thus, the attention measure does not qualify as an early warning device. On the other hand, Chinese Baidu web searches and compounded Baidu-Google queries predict BTC price dynamics at relatively high frequencies corresponding to fluctuations with periodicities lower than four to five months. The reaction time at relevant frequencies amounts to about two to four months. A rationalization of this relationship can be seen in momentum strategies (Jegadeesh and Titman, 2001) or endogenously timed herding (Süssmuth, 2002) rooted in short-term motives rather than fundamentals.

Irrespective of this result, I find that the cryptocurrency price is predictive for queries statistics across nearly all frequencies with a delay of just about one month.

Getting back to the overarching question of the economic forecasting potential and relevance of internet data raised in the introduction, one may summarize the central insights as follows. The BTC is not suited as collateral or as an element of a backing pool of securities in the context of asset-backed securities as its price dynamics shows bubbles for which (internet-based) popularity measures cannot act as early warning device. However, internet data from secondary sources can act quite well in the short and medium term (i.e. for within four to five months periods) as predictors for BTC price dynamics. This suggests to consider the BTC, for example, as a hedge candidate in periods of up to half-year intervals. At the same time, it is of paramount importance that popularity measures are carefully chosen so that the regional origin and distribution of BTC-related activity is taken care for.

In sum, I conclude that for the relationship between BTC price and web search query dynamics, the price clearly drives the hype. It does so virtually in real time or with a short delay of about one month. However, the hype also drives the price but only for relatively high frequency dynamics and with a reaction time of about one quarter.

References

Aalborg, H.A., Molnár, P., and J.E. de Vries (2019), What can explain the price, volatility and trading volume of Bitcoin?, forthcoming in *Finance Research Letters*

Alvarez-Ramirez, J., Rodriguez, E., and C. Ibarra-Valdez (2018), Long-range correlations and asymmetry in the Bitcoin market, *Physica A* 492, 948–955.

Andolfatto, D. and A. Spewak (2019), Whither the price of Bitcoin?, *Economic Synopses* 1, Federal Reserve Bank of St. Louis.

Baillie, R.T. and T. Bollerslev (1989), The message in daily exchange rates: a conditional-variance tale, *Journal of Business and Economic Statistics* 7, 297–305.

Bariviera, A.F., Basgall, M.J., Hasperué, W., and M. Naiouf (2017), Some styilzed facts of the Bitcoin market, *Physica A* 484, 82–90.

Böhme, R., Christin, N., Edelman, B., and T. Moore (2015), Bitcoin: economics, technology, and governance, *Journal of Economic Perspectives* 29, 18–35.

Brandvold, M., Molnár, P., Vagstad, K., and O.C.A. Vagstad (2015), Price discovery on Bitcoin exchanges, *Journal of International Financial Markets, Institutions & Money* 36, 18–35.

Breitung, J. and B. Candelon (2006), Testing for short- and long-run causality: a frequency domain approach, *Journal of Econometrics* 132, 363–378.

Breitung, J. and S. Schreiber (2018), Assessing causality and delay within a frequency band, *Econometrics and Statistics* 6, 57–73.

Cheah, E.-T. and J. Fry (2015), Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin, *Economics Letters* 130, 32–36.

Cheung, A. (W.-K.), Roca, E., and J.-J. Su (2015), Crypto-currency bubbles: an application of the Phillips–Shi–Yu (2013) methodology on Mt. Gox Bitcoin prices, *Applied Economics* 47, 2348–2358.

Ciaian, P., Rajcaniova, M., and d'A. Kancs (2016), The economics of BitCoin price formation, *Applied Economics* 48, 1799–1815.

Ciaian, P., Rajcaniova, M., and d'A. Kancs (2018), Virtual relationships: short- and long-run evidence from BitCoin and altcoin markets, *Journal of International Financial Markets, Institutions & Money* 52, 173–195.

Dolado, J.J. and H. Lütkepohl (1996), Making Wald tests work for cointegrated VAR systems, *Econometric Reviews* 15, 369–386.

Drost, F.C. and T.E. Nijman (1993), Temporal aggregation of GARCH processes, *Econometrica* 61, 909–927.

Edelman, B. (2012), Using internet data for economic research, *Journal of Economic Perspectives* 26, 189–206.

Foley, S., Karlsen, J.R., and T.J. Putņinš (2019), Sex, drugs, and Bitcoin: How much illegal activity is financed through cryptocurrencies?, forthcoming in *Review of Financial Studies*

Garcia, D., Tessone, C.J., Mavrodiev, P., and N. Perony (2014), The digital traces of bubbles: feedback cycles between socio-economic signals in the Bitcoin economy, *Journal of the Royal Society Interface* 11, 1–9.

Geweke, J. (1982), Measurement of linear dependence and feedback between multiple time series, *Journal of the American Statistical Association* 74, 427–431.

Gómez-González, J.E., Ojeda-Joya, J.N., Zárate-Solano, H., and F. Tenjo-Galarza (2014), Testing for causality between credit and real business cycles in the frequency domain: an illustration, *Applied Economics Letters* 21, 697–701.

Hafner, C.M. (2008), Temporal aggregation of multivariate GARCH processes, *Journal of Econometrics* 142, 467–483.

Hafner, C.M. (2019), Testing for bubbles in cryptocurrencies with time-varying volatility, forthcoming in *Journal of Financial Econometrics*

Hayes, A.S. (2019), Bitcoin price and its marginal cost of production: support for a fundamental value, *Applied Economics Letters* 26, 554–560.

He, G., Chen, Y., Chen, B., Wang, H., Shen, L., Liu, L., Suolang, D., Zhang, B., Ju, G., Zhang, L., Du, S., Jiang, X., Pan, Y., and Z. Min (2018), Using the Baidu Search Index to predict the incidence of HIV/AIDS, *Nature* (*Scientific Reports*) 8, 1–10.

Heer, B. and B. Süssmuth (2013), Tax bracket creep and its effects on income distribution, *Journal of Macroeconomics* 38, 393–408.

Homm, U. and J. Breitung (2012), Testing for speculative bubbles in stock markets: a comparison of alternative methods, *Journal of Financial Econometrics* 10, 198–231.

Jegadeesh, N. and S. Titman (2001), Profitability of momentum strategies: an evaluation of alternative explanations, *Journal of Finance* 56, 699–720.

Jian, X. and B.-S. Lee (2007), Stock returns, dividend yield, and book-to-market ratio, *Journal of Banking & Finance* 31, 455–475.

Kjærland, F., Khazal, A., Krogstad, E.A., Nordstrøm, F.B.G., and A. Oust (2018), An analysis of Bitcoin's price dynamics, *Journal of Risk and Financial Management* 11, 63–81.

Kristoufek, L. (2013), *BitCoin* meets *Google Trends* and *Wikipedia*: quantifying the relationship between phenomena of the internet era, *Nature* (*Scientific Reports*) 3, 1–7.

Kristoufek, L. (2015), What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis, *PLOS ONE* 10, 1–15.

Li, X. and C.A. Wang (2016), The technology and economic determinants of cryptocurrency exchange rates: the case of Bitcoin, *Decision Support Systems* 95, 49–60.

Li, Z.-Z., Tao, R., Su, C.W., and O.-R. Lobonţ (2019), Does Bitcoin bubble burst?, forthcoming in *Quality & Quantity*

Linnemer, L. (2012), Dissipative advertising signals quality: static model with a continuum of types, *Economics Letters* 114, 150–153.

Liu, K., Wang, T., Yang, Z., Huang, X., Milinovich, G.J., Lu, Y., Jing, Q., Xia, Y., Zhao, Z., Yang, Y., Tong, S., Hu, W., and J. Lu (2016), Using Baidu search index to predict Dengue outbreak in China, *Nature* (*Scientific Reports*) 6, 1–9.

Loughran, T. and B. McDonald (2011), When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks, *Journal of Finance* 66, 35–65.

Manuelli, R.E. and J. Peck (1990), Exchange rate volatility in an equilibrium asset pricing model, *International Economic Review* 31, 83–108.

Mora, C., Rollins, R.L., Taladay, K., Kantar, M.B., Chock, M.K., Shimada, M., and E.C. Franklin (2018), Bitcoin emissions alone could push global warming above 2°C, *Nature* (*Climate Change*) 8, 931–933.

Phillips, P.C.B., Wu, Y., and J. Yu (2011), Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values?, *International Economic Review* 52, 201–226.

Su, C.-W., Li, Z.-Z., Tao, R., and D.-K. Si (2018), Testing for multiple bubbles in bitcoin markets: A generalized sup ADF test, *Japan and the World Economy* 46, 56–63.

Süssmuth, B. (2002), Can herding explain cyclic comovement?, *International Game Theory Review* 4, 33–51.

Tastan, H. (2015), Testing for spectral Granger causality, *Stata Journal* 15, 1157–1166.

Toda, H.Y. and T. Yamamoto (1995), Statistical inference in vector autoregressions with possibly integrated processes, *Journal of Econometrics* 66, 225–250.

Yermack, D. (2015), Is Bitcoin a real currency? An economic appraisal, in Lee, D.K.C. (ed.), *The Handbook of Digital Currency* Part 1, Chapter 2, Academic Press Elsevier, Amsterdam, 31–43.

Yermack, D. (2017), Corporate governance and blockchains, *Review of Finance* 21, 7–31.

Appendix

A.1 Summary statistics of central series

Table A.1: Descriptive statistics of central time series (Section2.1; Section 3.1)

	Mean	Std. dev.	Min	Max	N obs	Range
BTC/USD	1182.98	2723.23	2.68	15065.28	83	05/2011–03/2018
	979.41	2420.78	2.68	15065.28	81	05/2011–01/2018
EXCHUS	6.39	0.24	6.05	6.92	81	05/2011–01/2018
BTC/CNY	6418.4	15865.47	17.05	99328.39	81	05/2011–01/2018
GTS	6.92	13.98	0	100	83	05/2011–03/2018
	6.31	13.53	0	100	81	05/2011–01/2018
BSI	414064.6	597153.6	3686	2794942	81	05/2011–01/2018
BSI (norm.'ed)	14.81	21.37	0.13	100	81	05/2011–01/2018
B-G-C	0.106	0.153	0.001	0.820	81	05/2011–01/2018

A.2 Software and Chinese data retrieval

To carry out tests and estimations I partly relied on code and software written by Benjamin Loeper, Hüseyin Tastan, and Sven Schreiber. Command suite `bcgcausality` for Stata and function package `delayspectral.gfn` for open-source econometrics program Gretl have been used. Chinese Baidu Search Index (BSI) data were retrieved on 19 February 2018 (in daily frequency), aggregated to monthly frequency, and made available to me by Jingjing Lyu. The usual disclaimer applies.