Does market attention affect Bitcoin returns and volatility?

Gianna Figà-Talamanca[a] and Marco Patacca[b]

[a]Department of Economics, University of Perugia, via A. Pascoli 20, 06123 Perugia, Italy; Department of Finance and Risk Engineering, Tandon School of Engineering, New York University, U.S.A. (Visiting Professor, 2017/2018); [b]Department of Economics, University of Perugia, via A. Pascoli 20, 06123 Perugia, Italy.

ARTICLE HISTORY
Compiled May 25, 2018

ABSTRACT
In this paper we measure market attention by applying several filters on time series for the trading volume or the SVI Google searches index. We analyze relative impact of these measures either on the mean or on the variance of Bitcoin returns by fitting non linear econometric models to historical data from January 1, 2012 to December 31, 2017; two non-overlapping subsamples are also considered. Outcomes confirm that market attention has an impact on Bitcoin returns. Specifically, trading volume related measures affect both the mean and the conditional variance of Bitcoin returns while internet searches volume mainly affects the conditional variance of returns.

KEYWORDS
Bitcoin; Market Attention; SVI Google Index; ARMA; GARCH

1. Introduction

Bitcoin[1] is a digital currency, built on a peer-to peer network and on the blockchain, a public ledger where all transactions are recorded and made available to all nodes. Opposite to traditional banking transactions, based on trust for counter-party, BitCoin relies on cryptography and on a consensus protocol for the network. The entire system is founded on an open source software created in 2009 by a computer scientist known under the pseudonym Satoshi Nakamoto, whose identity is still unknown. Hence, Bitcoin is an independent digital currency, not subject to the control of central authority and without inflation; furthermore, transactions in the network are pseudonymous and irreversible.

BitCoin and the underlying blockchain technology have gained much attention in the last few years. Research on BitCoin often deals with cybersecurity and legitimacy issues such as the analysis of double spending possibilities and other cyber-threats; recently, high returns and volatility have attracted research towards the analysis of Bitcoin price efficiency, such as Almudhaf (2018); Urquhart (2016); Nadarajah and Chu (2017), as well as its price dynamics. Within the latter branch of research a non-exhaustive list is Kristoufek (2013, 2015); Bukovina and Martiček (2016); Dyhrberg (2016); Ciaian, Rajcaniova, and Kancs (2016); Katsiampa (2017); Cretarola, Figà-Talamanca, and Patacca

CONTACT Gianna Figà-Talamanca. Email: gianna.figatalamanca@unipg.it, gft220@nyu.edu.

[1]We use the following rule throughout the paper: the term BitCoin refers to the whole system network while Bitcoin refers to the digital currency.

(2017); Blau (2017). Among quoted papers, many contributions claim that Bitcoin price is driven by attention or sentiment about the BitCoin system itself. Possible driving factors are Google searches, Wikipedia requests (Kristoufek 2013) or more traditional indicators as the number or volume of transactions (Kristoufek 2015). In Bukovina and Martiček (2016) sentiment data are obtained from `http://sentdex.com/`, a online platform specialized on natural language processing algorithms to deliver a positive, neutral or negative feeling about a specific topic. The dependence of Bitcoin price on investors' attention is also investigated in Ciaian, Rajcaniova, and Kancs (2016) where the authors analyse the dependence of Bitcoin price on several market forces jointly: supply and demand for Bitcoins, some variables related to global macroeconomic and financial development such as Stock market indices and oil price, and several attractiveness factors. Specifically, they measure attractiveness of Bitcoin by means of the number of Wikipedia inquiries on the topic, the number of new users and the number of posts in the online forum `https://bitcointalk.org/`. By estimating Vector AutoRegressive and Vector Error Correction models, they find that such variables are significant in explaing Bitcoin prices. As for more traditional attention measures note that in Blau (2017) a time series model is introduced in order to identify the dynamic relation between speculation activity and price; Bitcoin returns are regressed against a demeaned measure of trading activity, following the idea in Llorente et al. (2002) and regression errors are modeled as a standard GARCH(1,1) process to account for heteroschedasticity. Models within the GARCH family have also been applied to describe the dynamics of Bitcoin returns and volatility in Dyhrberg (2016); Katsiampa (2017) but neither attention nor sentiment are taken into account in their setting. Differently from previous contributions, we investigate whether and to which extent market attention influences the dynamics of Bitcoin either in the mean returns or in its volatility. To this end we measure Bitcoin attractiveness either by a classical measure of attention such as the total trading volume in the market or, as suggested in Da, Engelberg, and Gao (2011), by the Search Volume Index (SVI) provided by Google; the latter is particularly suitable in this framework since Bitcoin is an internet based digital currency and internet users commonly collect information through a search engine such as Google. With a different goal, Google trends data are also used by Yelowitz and Wilson (2015) to distinguish the characteristic of Bitcoin users. It is worth noticing that Urquhart (2018) investigates the motivations for Bitcoin attention, with a complementary to our approach; further comparisons will be given in concluding remarks. In Da, Engelberg, and Gao (2011) the authors also find strong evidence that SVI captures the attention of retail investors: "the search volume is likely to be representative of the internet search behavior of the general population and more critically, search is a revealed attention measure: if you search for a stock in Google, you are undoubtedly paying attention to it. Therefore, aggregate search frequency in Google is a direct and unambiguous measure of attention".

Indeed, we believe that many of the retail investors in Bitcoin, especially after its steady increase of value, enter the market for speculation purposes and their positions in the Bitcoin and cryptocurrency market depend heavily on news on the media, tweets of well-known investors or experts; their information on Bitcoin characteristics may be based and fed by performing internet searches as argued in Da, Engelberg, and Gao (2011). Such investors are responsible for noisy behavior of Bitcoin and may have strongly contributed to increase it volatility over time.

In order to test for the impact of attention on Bitcoin returns we estimate several time series models where the trading volume and the Google SVI Index (suitably filtered) are taken as exogenous factors. Overall, we find evidence that Bitcoin returns are driven by market attention and, within this framework we are able to assess best candidate models for the analyzed data-sets. In particular, the trading volume affects both the mean of Bitcoin returns and their volatility. The SVI index is strongly significant in the conditional variance of returns for all the time series under analysis while it is strongly significant in the mean equation only for the second subsample. The latter results is complementary with Ciaian, Rajcaniova, and Kancs (2016) where they found that views on Wikipedia affect the Bitcoin price only in the first period of its history. The authors motivate this finding by arguing that the initial investors needed to understand the underlying structure of the cryptocurrency and the online encyclopedia was the first source of information in order to gather some *knowledge* on Bitcoin. We agree with their view concerning the first period and we claim that in the last few years retail investors have become more interested in *news* on the web about Bitcoin rather than structural knowledge; they resort to a search engine such as Google to get these media news motivating the strong significance of Google searches in Bitcoin returns. The rest of the paper is structured as follows. In Section 2 we describe the alternative models for the Bitcoin price dynamics, in section 3 we present the methodology, in Section 4 we sum up the empirical findings and in Sections 5 we give some concluding remarks and draw directions for future investigations.

2. Bitcoin price modeling

2.1. *Data*

We consider daily data for the average price of Bitcoin across main exchanges, obtained by `https://blockchain.info/`, from January 1, 2012 to December 31, 2017. In Figures 1 we plot Bitcoin prices and returns and in Table 1 the corresponding descriptive statistics.

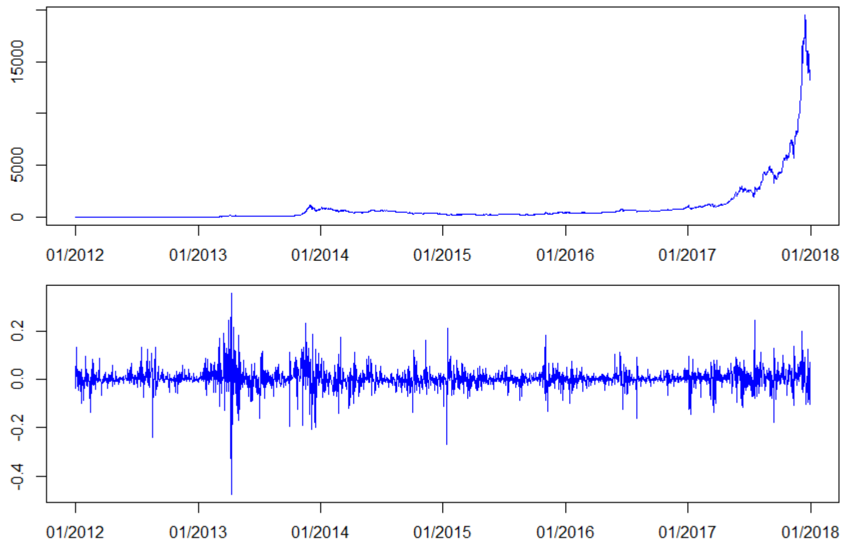

Figure 1.: Bitcoin data from January 1, 2012 to December 31, 2017: price (top) and logarithmic returns (bottom).

In order to account for time variability on outcomes we also split the available data in two subsamples: from January 1, 2012 to December 31, 2014 and from January 1, 2015 to December 31, 2017.

Table 1.: Summary Statistics of daily returns

	Whole sample	1st subsample	2nd subsample
Min.	-0.4783	-0.4783	-0.2686
Q_1	-0.0101	-0.0130	-0.0079
Median	0.0020	0.0011	0.0027
Mean	0.0036	0.0038	0.0035
Q_3	0.0180	0.0186	0.0179
Max.	0.3590	0.3590	0.2466
Standard Dev.	0.0457	0.0525	0.0377
Skewness	-0.7615	-0.9469	-0.1688
Kurtosis	20.2346	20.3505	10.9297
JB-Test p-value	0.0000	0.0000	0.0000

Both descriptive statistics and Jarque-Bera test p-values, reported in Table 1, evidence strong non-normality of returns across the whole time-series as well as the two subsamples.

In Figures 2, 3 we plot the auto-correlation and the partial auto-correlation functions of the Bitcoin logarithmic returns for the three time-series respectively; a significant serial dependence structure is evidenced both in the whole time series and in the first period under investigation, while it is reduced in the last time interval.

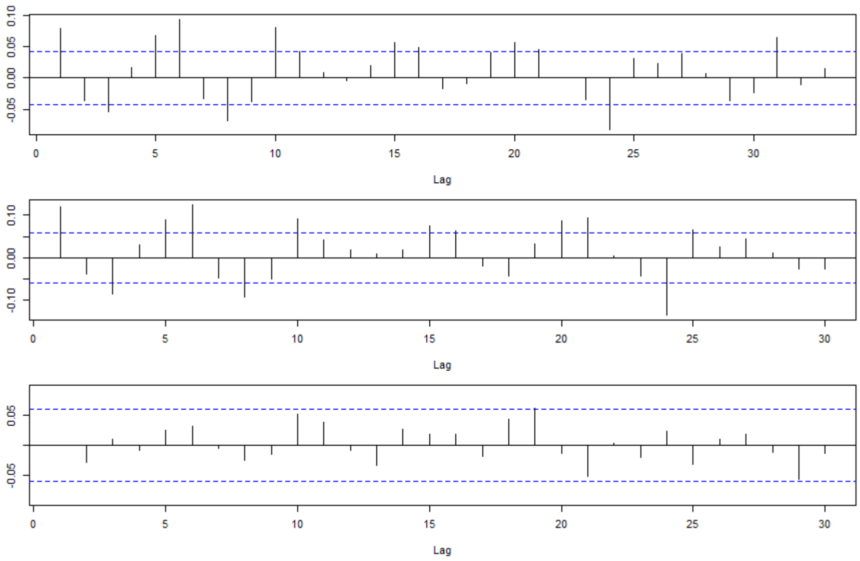

Figure 2.: Autocorrelation function for Bitcoin logarithmic returns: whole series (top), subsample 1 (center), subsample 2 (bottom)

2.2. *Models*

The serial dependence evidenced in Figures 2 and 3 suggests that the dynamics of Bitcoin returns may be described within the autoregressive moving average (ARMA) models. Since we are interested in the effect of market attention on Bitcoin returns we add an exogenous process in the model specification, denoted ARMA-X. For the sake of parameter parsimony we estimate an ARMA(1,1)-X to start with and move to higher number of lagged variables when necessary.

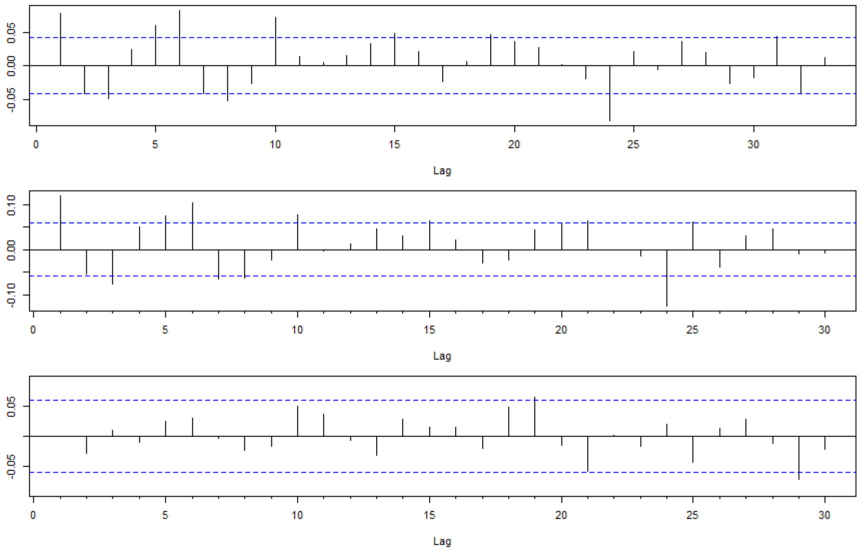

Figure 3.: Partial autocorrelation function for Bitcoin logarithmic returns: whole series (top), subsample 1 (center), subsample 2 (bottom)

The model specification for the ARMA(1,1)-X is:

$$R_t = a_0 + a_1 R_{t-1} + b_1 \epsilon_{t-1} + cX_t + \epsilon_t \tag{1}$$

where $\epsilon = \{\epsilon_t, t \geq 0\}$ is the error process, $X = \{X_t, t \geq 0\}$ is the attention measure and a_0, a_1, b_1, c are model parameters.

Figure 1 clearly show heteroschedasticity of returns in all the analyzed time periods. In order to take into account this feature the error process $\epsilon = \{\epsilon_t, t \geq 0\}$ in (1) is modeled within the GARCH family. Again, we fit a GARCH-X model on the Bitcoin returns time series, including an exogenous variable representing market attention also in the conditional variance equation. Several model specifications are available within this framework; we believe that outcomes will not differ substantially from a qualitative viewpoint so we focus on two examples, the standard GARCH and the Exponential GARCH models; for the sake of parameter parsimony we start with the simple GARCH(1,1)-X and EGARCH(1,1)-X as possible specifications for the conditional variance.

Summing up, we describe Bitcoin returns with (1) and assume that $\epsilon_t = \sqrt{h_t}\eta_t$, where $\eta = \{\eta_t, t \geq 0\}$ is a Gaussian noise and

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1} + \gamma X_t \tag{2}$$

or

$$\log h_t = \alpha_0 + \alpha_1 \eta_{t-1} + \beta_1 \log h_{t-1} + \lambda \left(|\eta_{t-1}| - \mathbb{E}\left[|\eta_{t-1}|\right]\right) + \gamma X_t \tag{3}$$

for the GARCH and EGARCH model, respectively.

We also estimate nested models by setting some of parameters to zero as well as a standard regression model (LR) on the attention process as a benchmark.

3. Methodology

3.1. *Market attention variables*

The exogenous variables representing market attention are based on two sources of data: the total volume of transactions in Bitcoins, provided by `https://blockchain.info/`, and the adjusted volume of internet searches, the Google SVI index, delivered by `https://trends.google.com/trends/`. Denoting with A the available time series, we consider as alternative measures of attention the variables $X_1 := \log(A)$, $X_2 := \Delta \log(A)$ and $X_3 := |X_2|$ obtained by suitably filtering the raw series A. The logarithm filter is applied in order to have a scale reduction: volume traded and volume searches have very high values with respect to returns so, if not scaled, the estimated coefficients in (1),(2) and (3) would be negligible, though statistically significant. The differenced variable is considered to understand whether the variations affects Bitcoin more significantly than the attention level; finally the third variable is accounted for in order to investigate if either the magnitude or the sign is more likely to affect Bitcoin returns.

In Figure 4 we plot the trading volume and the SVI index, both in the logarithmic scale, and in Table 2 we sum up the corresponding descriptive statistics for three samples. In order to check for stationarity we perform the Dickey Fuller test for both attention measures; the p-values of the tests are reported in last row of Table 2. It is worth noticing that the volume of transactions is stationary with a non-zero mean, while the SVI index is stationary around a deterministic trend.

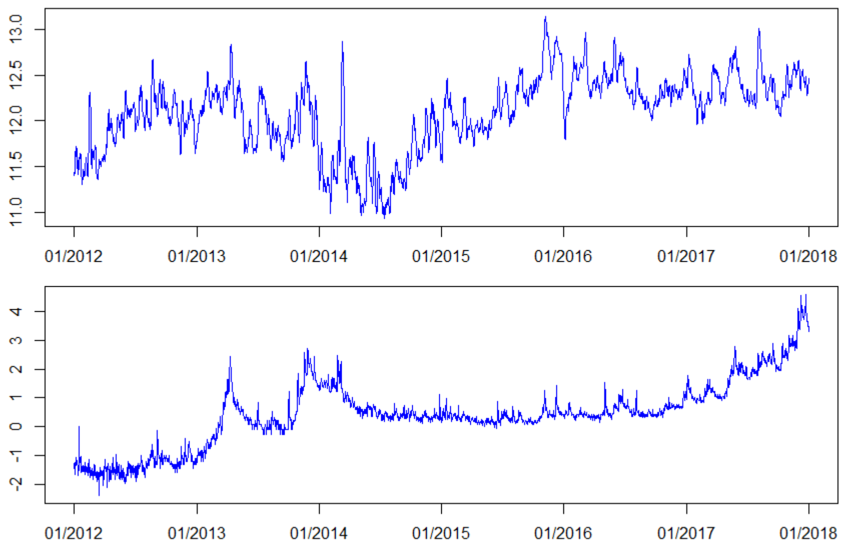

Figure 4.: Bitcoin trading Volume (top) and Google Searches Volume Index (bottom) observed from January 1, 2012 to December 31, 2017 in logarithmic scale

3.2. *Model selection procedure*

In what follows we estimate model (1), (2) and (3) as well as nested models where the attention variable X_t is replaced alternatively by X_1, X_2 and X_3. Best models are selected according to the Akaike and the Bayesian Information Criteria (AIC, BIC). We also test for residual serial dependence and heteroschedasticity for the selected models; it is well known that if a model specification is suitable to describe the dynamics of

Table 2.: Summary Statistics of Attention variables

	Whole sample		1st subsample		2nd subsample	
	Vol.	SVI	Vol.	SVI	Vol.	SVI
Min.	10.933	-2.401	10.933	-2.401	11.549	-0.048
Q_1	11.848	0.092	11.564	-1.197	12.184	0.344
Median	12.165	0.403	11.879	0.195	12.327	0.526
Mean	12.088	0.430	11.842	-0.090	12.334	0.951
Q_3	12.364	0.945	12.140	0.590	12.490	1.227
Max.	13.143	4.605	12.871	2.730	13.143	4.605
Standard Dev.	0.409	1.122	0.387	1.075	0.254	0.907
Skewness	-0.525	0.255	-0.187	0.082	0.117	1.532
Kurtosis	2.911	3.640	2.356	2.088	3.137	4.793
Dickey-Fuller p-value	0.000	0.000	0.005	0.002	0.002	0.005

returns, then the residual process should be a white noise. Hence, we test the null hypothesis of uncorrelated residuals by means of the Ljung-Box Q-Test, see Ljung and Box (1978), and the null of homoschedasticity via the Engle's ARCH Test, see Engle (1982); the Ljung-Box Q^2-Test, applied to the squared residuals, is useful to detect serial non-linear dependence. Tests are performed for several choices of the maximum number of Lags.

The fitting procedure is performed step by step. At first we make sure that a simple linear model such as ARMA-X is not suitable to describe Bitcoin returns. Indeed, residuals of all ARMA specifications, considered according to the different filtered measures of attention, still exhibit heteroschedasticity, for both cases of trading volume and Google SVI index. The empirical results of this preliminary exercise are not reported in this paper but are available under request. As a second step we estimate the full specification in (1) with either (2) or (3) by including the same exogenous variable in both the mean and the variance equations and select the best performing models in terms of the AIC and BIC values. All outcomes are reported in panel (a) of following Tables. Then, in order to have further insights on the relevance of various filters, we fit models allowing for different filtered variables in the mean and in the variance equations; the outcomes are summed up in panel (b) of following tables, only for specifications with similar or better values of the AIC and BIC with respect to previous cases. Results for all crossed models specifications are available upon request. Panel (c) in all of the Tables includes results for higher ARMA order specification in the mean, when these are necessary to remove residual serial correlation of the returns.

4. Empirical Results

4.1. *Empirical results using daily traded Volume*

In Table 3 panel (a) we sum up the Akaike Information and Bayesian Information Criteria for all competitor models when considering the same filtered variable both in the mean and in the variance equation. The best models within this setting, underlined in Table 3, are the LR-X_2 and ARMA(1,1)-X_2 with EGARCH(1,1)-X_2 errors. For selected models, the exogenous variable (log-difference of trading volume) is not or slightly significant in the mean equation while it is strongly significant in the variance equation. Besides, the added ARMA parameters in the mean do not increase the absolute value of the two information criteria, hence, if one seeks for parameter parsimony the best overall choice would be the simple linear regression specification.

Looking further into the outcomes, it is evident that the log filtered variable X_1 is

Table 3.: AIC and BIC model selection analysis (whole series, trading volume case)

Model		AIC	BIC	X_m	X_v
Panel a: models with same filtered variables					
LR	GARCH(1,1)	-8488.59	-8465.80	-	-
LR-X_1	GARCH(1,1)-X_1	-8493.41	-8459.23	**	
LR-X_2	GARCH(1,1)-X_2	-8578.86	-8544.90	*	****
LR-X_3	GARCH(1,1)-X_3	-8550.82	-8516.64		****
LR	EGARCH(1,1)	-8507.43	-8479.17	-	-
LR-X_1	EGARCH(1,1)-X_1	-8515.98	-8476.10	****	
<u>LR-X_2</u>	<u>EGARCH(1,1)-X_2</u>	<u>-8633.20</u>	<u>-8593.32</u>		****
LR-X_3	EGARCH(1,1)-X_3	-8542.71	-8502.83	*	****
ARMA(1,1)	GARCH(1,1)	-8491.22	-8457.26	-	-
ARMA(1,1)-X_1	GARCH(1,1)-X_1	-8495.38	-8449.81	**	
ARMA(1,1)-X_2	GARCH(1,1)-X_2	-8581.71	-8536.36		****
ARMA(1,1)-X_3	GARCH(1,1)-X_3	-8552.57	-8507.00		
ARMA(1,1)	EGARCH(1,1)	-8510.50	-8470.63	-	-
ARMA(1,1)-X_1	EGARCH(1,1)-X_1	-8519.27	-8468.00	****	
<u>ARMA(1,1)-X_2</u>	<u>EGARCH(1,1)-X_2</u>	<u>-8635.17</u>	<u>-8584.12</u>	*	****
ARMA(1,1)-X_3	EGARCH(1,1)-X_3	-8544.68	-8493.41	**	****
Panel b: models with different filtered variables					
LR	EGARCH(1,1)-X_2	-8634.29	-8600.11	-	****
ARMA(1,1)	EGARCH(1,1)-X_2	-8636.26	-8590.69	-	****
LR-X_1	**EGARCH(1,1)-X_2**	**-8647.00**	**-8607.12**	********	********
ARMA(1,1)-X_1	**EGARCH(1,1)-X_2**	**-8648.10**	**-8596.83**	********	********
Panel c: models with higher ARMA order					
AR(6)[1]-X_1	**EGARCH-(1,1)X_2**	**-8648.32**	**-8597.05**	********	********
ARMA(6,1)[2]-X_1	EGARCH(1,1)-X_2	-8647.00	-8590.03	****	****

Columns from left to right represent the mean equation model, the variance equation model, the AIC, the BIC, the significance of the explanatory variable in the mean X_m, the significance of the explanatory variable in the variance X_v.
* $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; **** $P \leq 0.0001$.
[1] AR(6)-X with parameters $a_2 = a_3 = a_4 = a_5 = 0$.
[2] ARMA(6,1)-X with parameters $a_2 = a_3 = a_4 = a_5 = 0$

significant in the mean for nearly all competing models, but is not significant in the variance; the opposite for the differences X_2. It is worth to investigate whether a model specification with X_1 in the mean and X_2 in the variance equation is capable to give a better fit. Indeed, if we allow for crossed variables in mean and variance equations, the best overall models, denoted in bold in Table 3, are the LR-X_1 EGARCH(1,1)-X_2 and the ARMA(1,1)-X_1 EGARCH(1,1)-X_2 and both explanatory variables are strongly significant. Again, if parsimony of parameters is important the former is the best choice.

Outcomes for diagnostic tests of selected best models are shown in Table 4: note that the residuals still exhibit serial correlation for the LR-X_1 EGARCH(1,1)-X_2 case which is partially removed by augmenting the lags order, via an ARMA(1,1)-X_1 EGARCH(1,1)-X_2. In order to remove serial correlation we need to add a 6 days lagged values i.e. an AR(6)[1]-X_1 EGARCH(1,1)-X_2. The AIC and BIC values for this case are reported in panel (c) of Table 3. Again the two explanatory variables are strongly significant.

Table 4.: Diagnostics for competing non-linear models (whole series, trading volume case)

Test		p-value				
		lag=1	lag=5	lag=10	lag=15	lag=20
LR-X_1 EGARCH(1,1)-X_2	Ljung-Box Q	0.0001	0.0006	0.0002	0.0001	0.0001
	Ljung-Box Q^2	-	0.0052	0.1176	0.3588	0.6294
	Engle's Arch	0.0090	0.0378	0.3030	0.5920	0.8102
ARMA(1,1)-X_1 EGARCH(1,1)-X_2	Ljung-Box Q	-	0.0170	0.0039	0.0029	0.0024
	Ljung-Box Q^2	-	0.0071	0.1422	0.4030	0.6832
	Engle's Arch	0.0117	0.0475	0.3421	0.6368	0.8469
AR(6)[1]-X_1 EGARCH(1,1)-X_2	Ljung-Box Q	-	0.0217	0.0343	0.0256	0.0200
	Ljung-Box Q^2	-	0.0057	0.1218	0.3615	0.6452
	Engle's Arch	0.0084	0.0507	0.3408	0.6347	0.8536

[1] AR(6)-X with parameters $a_2 = a_3 = a_4 = a_5 = 0$.

As we can see in Table 4, in all of the three competing models the residuals are rejected to be heteroschedastic. Since the gain in explaining serial correlation is not much across the three models and might worsen the BIC values, we believe that the overall best is the simple LR-X_1 EGARCH(1,1)-X_2 model.

In Table 5 we summarize the estimated parameters for the chosen specification; note that the only non significant parameter is the coefficient of the standardized error in the variance equation, namely α_1.

Table 5.: Parameter estimates for the LR-X_1 EGARCH(1,1)-X_2 model (whole series, trading volume case)

	Estimate	Std. Error	t value	Pr(> $\|t\|$)
a_0	-0.061991	0.000854	-72.5654	0.00000
c	0.005296	0.000071	74.1544	0.00000
α_0	-0.173573	0.031312	-5.5434	0.00000
α_1	0.018756	0.010924	1.7169	0.08599
β_1	0.970030	0.004689	206.8636	0.00000
λ	0.263746	0.020464	12.8885	0.00000
γ	2.380453	0.195636	12.1678	0.00000

Summing up, the trading volume is strongly significant both in the mean and in the variance of Bitcoin returns; in particular, the level of the trading volume is strongly significant in the mean while its changes are strongly significant in the variance equation.

[1] AR(6)-X with parameters $a_2 = a_3 = a_4 = a_5 = 0$

Performing the same analysis separately for each subsample[2], we find similar results to the previous case in the first subsample, while in the second subsample the volume affects only the variance. Furthermore the best overall model is the AR(1)-X_1 EGARCH(1,1)-X_2 for the first period and the LR EGARCH(1,1)-X_2 for the second. In Table 6 we display the diagnostic tests p-values. A slightly different behavior is evident for the two time periods: in the latter the simple linear regression model is able to explain both serial dependence and heteroschedasticity in the data; instead in the former we need to choose the AR specification to obtain similar results.

Table 6.: AR(1)-X_1 EGARCH(1,1)-X_2 for the first subsample and LR EGARCH(1,1)-X_2 for the second (trading volume case)

Test		p-value				
		lag=1	lag=5	lag=10	lag=15	lag=20
1st subsample	Ljung-Box Q	-	0.5603	0.5020	0.4419	0.4438
	Ljung-Box Q^2	-	0.4131	0.7642	0.9492	0.9913
	Engle's Arch	0.3042	0.7411	0.9045	0.9846	0.9975
2nd subsample	Ljung-Box Q	0.0391	0.2067	0.0427	0.0640	0.1114
	Ljung-Box Q^2	-	0.0030	0.0471	0.1385	0.1507
	Engle's Arch	0.0034	0.0227	0.1238	0.4585	0.4748

4.2. *Empirical results using daily Google SVI index*

Let us estimate the suggested models on the time series of SVI Google Searches following the same steps as in previous section. Table 7 sums up the AIC and BIC outcomes as well as significance of market attention measures.

Again, the selection is controversial since the LR-X_2 EGARCH(1,1)-X_2 minimizes the BIC criterion while AIC indicates the AR(6)[3]-X_2 EGARCH(1,1)-X_2 as the best overall model. Note that market attention, measured by the number of Google searches or related filtered variables, is strongly significant in the variance equation in nearly all cases, while it is significant only for some model specifications in the mean term. More precisely both the mean and the variance of returns are affected by logarithmic differences of Google SVI index while the index level (raw or detrended) affects weakly and only in few competitor cases the mean equation.

Outcomes for Ljung-Box and Engle's ARCH tests are displayed in Table 8. Differently from the trading volume case, heteroschedasticity of returns is ruled out by both selected models while returns are still autocorrelated in the LR-X_2 EGARCH(1,1)-X_2. For this reason we believe that the overall best in this case is the AR(6)[3]-X_2 EGARCH(1,1)-X_2.

In Table 9 the estimated parameters for model AR(6)[3]-X_2 EGARCH(1,1)-X_2. Non significant parameters are α_1 and α_2, the coefficient of autoregressive of order 1 and 5 respectively. However without including these variable we are not able to remove serial correlation.

Overall, the Google SVI index strongly affects the variance of Bitcoin returns while it is only weakly significant in the mean equation; differently from the previous case the index changes are evidenced to mostly affect Bitcoin returns rather than the index level.

The analysis of the two subsamples delivers similar qualitative results and detailed outcomes are available upon request. For the first period model selection is towards

[2] Detailed outcomes are available upon request.

[3] AR(6)-X with parameters $a_2 = a_3 = a_4 = 0$

Table 7.: AIC and BIC model selection analysis (whole series, Google SVI index case)

Model		AIC	BIC	X_m	X_v
Panel a: models with same filtered variables					
LR	GARCH(1,1)	-8488.59	-8465.80	-	-
LR-X_1	GARCH(1,1)-X_1	-8495.38	-8461.20		****
LR-X_2	GARCH(1,1)-X_2	-8722.37	-8688.19	****	****
LR-X_3	GARCH(1,1)-X_3	-8584.78	-8550.60		
LR	EGARCH(1,1)	-8507.43	-8479.17	-	-
LR-X_1	EGARCH(1,1)-X_1	-8512.69	-8472.82		**
LR-X_2	**EGARCH(1,1)-X_2**	**-8933.80**	**-8893.93**	**	****
LR-X_3	EGARCH(1,1)-X_3	-8567.91	-8528.03		****
ARMA(1,1)	GARCH(1,1)	-8491.22	-8457.26	-	-
ARMA(1,1)-X_1	GARCH(1,1)-X_1	-8498.89	-8453.32		****
ARMA(1,1)-X_2	GARCH(1,1)-X_2	-8732.45	-8686.88	****	****
ARMA(1,1)-X_3	GARCH(1,1)-X_3	-8586.75	-8541.18		****
ARMA(1,1)	EGARCH(1,1)	-8510.50	-8470.63	-	-
ARMA(1,1)-X_1	EGARCH(1,1)-X_1	-8515.10	-8463.83	*	**
ARMA(1,1)-X_2	EGARCH(1,1)-X_2	-8932.93	-8881.88	***	****
ARMA(1,1)-X_3	EGARCH(1,1)-X_3	-8572.51	-8521.24		****
Panel b: models with different filtered variables					
LR	EGARCH(1,1)-X_2	-8925.26	-8891.08	-	****
ARMA(1,1)	EGARCH(1,1)-X_2	-8924.82	-8879.25	-	****
LR-X_1	EGARCH(1,1)-X_2	-8927.01	-8887.13	***	****
ARMA(1,1)-X_1	EGARCH(1,1)-X_2	-8925.70	-8874.65	***	****
Panel c: models with higher ARMA order					
AR(6)[1]-X_2	**EGARCH(1,1)-X_2**	**-8937.09**	**-8880.12**	****	****
ARMA(6,1)[2]-X_2	EGARCH(1,1)-X_2	-8935.56	-8872.89	****	****

Columns from left to right represent the mean equation model, the variance equation model, the AIC, the BIC, the significance of the explanatory variable in the mean X_m, the significance of the explanatory variable in the variance X_v.
* $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; **** $P \leq 0.0001$.
[1] AR(6)-X with parameters $a_2 = a_3 = a_4 = 0$.
[2] ARMA(6,1)-X with parameters $a_2 = a_3 = a_4 = 0$.

Table 8.: Diagnostics for competing non-linear models (whole series, Google SVI index case)

Test		p-value				
		lag=1	lag=5	lag=10	lag=15	lag=20
LR-X_2 EGARCH(1,1)-X_2	Ljung-Box Q	0.0004	0.0014	0.0004	0.0006	0.0005
	Ljung-Box Q^2	-	0.3397	0.3446	0.6982	0.9090
	Engle's Arch	0.7394	0.6530	0.5630	0.8250	0.9481
AR(6)[1]-X_2 EGARCH(1,1)-X_2	Ljung-Box Q	-	0.0095	0.0484	0.0723	0.0564
	Ljung-Box Q^2	-	0.3284	0.4332	0.7822	0.9449
	Engle's Arch	0.7943	0.6328	0.6467	0.8748	0.9648

[1] AR(6)-X with parameters $a_2 = a_3 = a_4 = 0$.

Table 9.: Parameter estimates for the AR(6)[1]-X_2 EGARCH(1,1)-X_2 model (whole series, Google SVI index case)

	Estimate	Std. Error	t value	Pr(> $\|t\|$)
a_0	0.001585	0.000453	3.5016	0.000462
a_1	0.029688	0.020891	1.4211	0.155287
a_5	0.023004	0.020752	1.1085	0.267636
a_6	0.048577	0.017025	2.8533	0.004327
c	0.010218	0.002471	4.1355	0.000035
α_0	-0.157260	0.024078	-6.5313	0.000000
α_1	-0.050700	0.011899	-4.2610	0.000020
β_1	0.974485	0.003479	280.0698	0.000000
λ	0.259720	0.020271	12.8121	0.000000
γ	2.934440	0.150161	19.5420	0.000000

[1] AR(6)-X with parameters $a_2 = a_3 = a_4 = 0$.

LR EGARCH(1,1)-X_2 while for the most recent time series the best choice is LR-X_1 EGARCH(1,1)-X_2. These results highlight that the Google SVI Index affects the variance of Bitcoin returns in the first subsample while both the mean and variance equations in more recent years.

Table 10 reports the p-values of Ljung-Box test on model residuals and square residuals and the Engle's ARCH test p-values. Similarly to the volume case we obtain different behaviors for the two periods: in the latter subsample the linear regression model is able to explain both serial dependence and heteroschedasticity in the data; instead, in the former, the residuals still exhibit serial correlation that we may remove using an ARMA(1,1) EGARCH(1,1)-X_2. However in the latter model the BIC values substantially worsen.

Table 10.: LR EGARCH(1,1)-X_2 for the first subsample and LR-X_1 EGARCH(1,1)-X_2 model for the second (Google SVI index case)

Test		p-value				
		lag=1	lag=5	lag=10	lag=15	lag=20
1st subsample	Ljung-Box Q	0.0001	0.0017	0.0035	0.0107	0.0116
	Ljung-Box Q^2	-	0.3158	0.1040	0.2488	0.3533
	Engle's Arch	0.1965	0.0682	0.1569	0.3825	0.3674
2nd subsample	Ljung-Box Q	0.0120	0.2006	0.4115	0.7125	0.7783
	Ljung-Box Q^2	-	0.0316	0.1175	0.1528	0.2347
	Engle's Arch	0.0034	0.0227	0.1238	0.4585	0.4748

4.3. *Empirical results using daily Volume and Google SVI index*

In this section we estimate the model in Section 2.2 using the volume of transactions and the Google SVI index jointly. We denote with X^{vol} and X^{SVI} the exogenous variables related to volume of transactions and Google SVI index respectively and with $\mathbf{X} = (X^{vol}, X^{SVI}), \mathbf{c} = (c_1, c_2), \gamma = (\gamma_1, \gamma_2)$ the corresponding vector. Table 11 sums up the AIC and BIC results as well as significance of market attention measures for the whole time series.

Outcomes in panel (a) clearly show that the Google SVI index does not affect the mean of Bitcoin returns, therefore we estimate models in panel (b) and (c) excluding X_m^{SVI}. Furthermore, volume of transactions is weakly significant in the variance equation for some model specifications, then we omit X_v^{vol} in these models in panel (b) and (c).

Table 11.: AIC and BIC model selection analysis (whole series, jointly case)

Model		AIC	BIC	$\mathbf{X_m}$		$\mathbf{X_v}$	
				X_m^{vol}	X_m^{SVI}	X_v^{vol}	X_v^{SVI}
Panel a: models with same filtered variables							
LR	GARCH(1,1)	-8488.59	-8465.80	-	-	-	-
LR-$\mathbf{X_1}$	GARCH(1,1)-$\mathbf{X_1}$	-8492.10	-8446.74	****			
LR-$\mathbf{X_1}$	GARCH(1,1)-$\mathbf{X_2}$	-8749.54	-8703.97	****	****	****	****
LR	EGARCH(1,1)	-8507.43	-8479.17	-	-	-	-
LR-$\mathbf{X_1}$	EGARCH(1,1)-$\mathbf{X_1}$	-8522.77	-8471.72	****		*	**
LR-$\mathbf{X_1}$	EGARCH(1,1)-$\mathbf{X_2}$	-8968.86	-8917.59	****		****	****
ARMA(1,1)	GARCH(1,1)	-8491.22	-8457.26	-	-	-	-
ARMA(1,1)-$\mathbf{X_1}$	GARCH(1,1)-$\mathbf{X_1}$	-8491.66	-8434.91	***			
ARMA(1,1)-$\mathbf{X_1}$	GARCH(1,1)-$\mathbf{X_2}$	-8745.60	-8688.85	****			****
ARMA(1,1)	EGARCH(1,1)	-8510.50	-8470.63	-	-	-	-
ARMA(1,1)-$\mathbf{X_1}$	EGARCH(1,1)-$\mathbf{X_1}$	-8525.62	-8462.96	****		*	**
ARMA(1,1)-$\mathbf{X_1}$	EGARCH(1,1)-$\mathbf{X_2}$	-8968.20	-8905.54	****		****	****
Panel b: models with different filtered variables							
LR-$\mathbf{X_1}$	EGARCH(1,1)-$\mathbf{X_2}$	-8935.56	-8895.68	****	-	-	****
ARMA(1,1)-$\mathbf{X_1}$	EGARCH(1,1)-$\mathbf{X_2}$	-8934.46	-8883.19	****	-	-	****
LR-$\mathbf{X_1}$	**EGARCH(1,1)-$\mathbf{X_2}$**	**-8970.61**	**-8925.04**	********	**-**	********	********
ARMA(1,1)-$\mathbf{X_1}$	EGARCH(1,1)-$\mathbf{X_2}$	-8969.95	-8912.99	****	-	****	****
Panel c: models with higher ARMA order							
AR(1)-$\mathbf{X_1}$	EGARCH(1,1)-$\mathbf{X_2}$	-8934.90	-8889.33	****	-	-	****
AR(6)[2]-$\mathbf{X_1}$	EGARCH(1,1)-$\mathbf{X_2}$	-8937.75	-8886.48	****	-	-	****
AR(1)-$\mathbf{X_1}$	EGARCH(1,1)-$\mathbf{X_2}$	-8970.39	-8919.12	****	-	****	****
AR(6)[1]-$\mathbf{X_1}$	**EGARCH(1,1)-$\mathbf{X_2}$**	**-8970.61**	**-8913.65**	********	**-**	********	********

Columns from left to right represent the mean equation model, the variance equation model, the AIC, the BIC, the significance of the explanatory variables in the mean X_m^{vol} and X_m^{SVI}, the significance of the explanatory variables in the variance X_v^{vol} and X_v^{SVI}.

* $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; **** $P \leq 0.0001$.

[1] AR(6)-X with parameters $a_2 = a_3 = a_4 = a_5 = 0$.

According to the AIC criterion we obtain the LR-X_1 EGARCH(1,1)-X_2 and AR(6)[4]-X_1 EGARCH(1,1)-X_2 as the best models while the BIC suggests the first one. It is important to note that in both cases SVI doesn't affect the conditional mean of Bitcoin returns while the volume of transactions is strongly significant in the mean as well as in variance.

Table 12 shows the diagnostic test for the selected best models. The AR(6)[4]-X_1 EGARCH(1,1)-X_2 model is the only able to explain the autocorrelation as well as the heteroschedasticity of returns, therefore we believe that it is the optimal choice.

Table 12.: Diagnostics for competing non-linear models (whole series, jointly case)

Test		p-value				
		lag=1	lag=5	lag=10	lag=15	lag=20
LR-X_1 EGARCH(1,1)-X_2	Ljung-Box Q	0.0007	0.0065	0.0046	0.0099	0.0094
	Ljung-Box Q^2	-	0.4475	0.3954	0.7746	0.9287
	Engle's Arch	0.9319	0.7415	0.5802	0.8541	0.9475
AR(6)[1]-X_1 EGARCH(1,1)-X_2	Ljung-Box Q	-	0.0210	0.0751	0.1293	0.1184
	Ljung-Box Q^2	-	0.4309	0.4217	0.7882	0.9389
	Engle's Arch	0.9940	0.7121	0.6029	0.8550	0.9502

[1] AR(6)-X with parameters $a_2 = a_3 = a_4 = a_5 = 0$.

In Table 13 estimated parameters for model AR(6)[4]-X_1 EGARCH(1,1)-X_2.

Table 13.: Parameter estimates for the AR(6)[1]-X_1 EGARCH(1,1)-X_2 model (whole series, jointly case)

	Estimate	Std. Error	t value	Pr(> $\|t\|$)
a_0	-0.053016	0.000860	-61.6380	0.000000
a_1	0.029409	0.016603	1.7713	0.076508
a_6	0.034362	0.019481	1.7639	0.077755
c_1	0.004520	0.000072	63.1500	0.000000
c_2	-	-	-	-
α_0	-0.135830	0.021943	-6.1902	0.000000
α_1	-0.042927	0.011890	-3.6105	0.000306
β_1	0.977956	0.003160	309.5078	0.000000
λ	0.238036	0.019059	12.4893	0.000000
γ_1	1.173072	0.198201	5.9186	0.000000
γ_2	2.652013	0.157923	16.7930	0.000000

[1] AR(6)-X with parameters $a_2 = a_3 = a_4 = a_5 = 0$.

Summing up, the trading volume significantly explains both mean returns and volatility in the Bitcoin market. In addition, investors whose attention is represented by Google searches volume do contribute significantly to an increase in Bitcoin volatility: retail investors use the Google search engine to look for news on which they may base their decision to enter or exit the Bitcoin market for speculative purposes, adding noise to its returns.

Performing the same analysis separately for each subsample, we find slightly different outcomes. In the first one the best overall model is AR(1)-X_1[5] EGARCH(1,1)-X_2 and, as in the whole series, the SVI affects only the variance of returns. Instead, the LR-X_1 EGARCH(1,1)-X_2 is the best model for the second one and in this case SVI as well as volume are strongly significant both in the variance and in the conditional mean. This feature is consistent with the above interpretation of retail investors and the SVI index. Indeed, in the last few years Bitcoin has been the object of huge media interest which

[4] AR-X(6) with parameters $a_2 = a_3 = a_4 = a_5 = 0$

may have boosted the number of noisy investors, thus affecting also Bitcoin returns. In Table 14 we display the diagnostic test results in this two cases.

Table 14.: AR(1)-X_1[1] EGARCH(1,1)-X_2 for the first subsample and LR-X_1 EGARCH(1,1)-X_2 model for the second (jointly case)

Test		p-value				
		lag=1	lag=5	lag=10	lag=15	lag=20
1st subsample	Ljung-Box Q	-	0.3459	0.4988	0.7624	0.7561
	Ljung-Box Q^2	-	0.3490	0.2604	0.5010	0.6715
	Engle's Arch	0.6183	0.6568	0.4624	0.6370	0.7482
2nd subsample	Ljung-Box Q	0.9247	0.9526	0.4651	0.4079	0.5838
	Ljung-Box Q^2	-	0.2344	0.0231	0.0670	0.1330
	Engle's Arch	0.5046	0.5125	0.0466	0.1065	0.1446

[1] In the mean term there is only X_m^{vol}

5. Concluding Remarks

The recent increasing trend in Bitcoin prices has pushed a new interest in the modeling of its returns. In Katsiampa (2017) the author compares several GARCH model specifications to model Bitcoin returns and volatility; in Dyhrberg (2016) a similar analysis is performed by adding financial risk factors such as Stock market indexes, fiat currency exchange rates and Gold spot and future prices to the mean equation. Many papers have suggested that Bitcoin price and returns are affected by market attention, see Kristoufek (2013, 2015), and sentiment, see Bukovina and Martiček (2016); we give further insights within this latter strand of literature by investigating whether such factors indeed influence Bitcoin price dynamics. More precisely, in this paper we estimate non-linear models where an attention-related exogenous variable is also included. Following the suggestions in Kristoufek (2013, 2015) we use either the SVI Google index or the trading volume of transactions to define market attention and we compute, by applying proper filters, several related variables; if A denotes the time series of trading volume or the SVI index, possible measures of attention are given by $X_1 := \log(A)$, $X_2 := \Delta \log(A)$ and $X_3 := |X_2|$. These variables are jointly or alternatively introduced as regressors both in the mean and in the conditional variance equations for Bitcoin returns, as defined in (1), (2) and (3).

For each specification we evaluate the Akaike and Bayesian Information Criteria and model validation is performed by applying classical diagnostic tests on the residuals. The analysis is conducted for three time series of Bitcoin returns: January 1, 2012 - December 31, 2014, January 1,2015 - December 31, 2017 and the whole data-set. The overall picture which can be drawn by our results is that attention measures affect significantly both the conditional mean and the conditional variance of Bitcoin returns in every model specification: further, it makes a positive contribution both in terms of AIC and BIC compared to the relative standard models. In particular, for the trading volume, the coefficients of market attention are significant for all specifications: the level of the volume trade X_1 is strongly significant in the mean equation whereas its changes X_2 and its absolute changes X_3 are strongly significant in the variance term. For the SVI index instead, the attention changes X_2 and X_3 are strongly significant in the variance term while the contribution on the mean returns is not significant but for few

[5] In the mean term there is only X_m^{vol}

cases, for the level X_1. This is consistent with our initial conjecture that investors whose attention is represented by Google SVI index do contribute significantly to an increase in Bitcoin volatility. Furthermore we show that the trading volume and the SVI index affect the Bitcoin returns even when taken jointly. This means that these measures are not redundant and add explanatory power if both included in the model. The larger values of information criteria in the jointly case confirm these findings. As remarked in the introduction, we address a complementary research question to Urquhart (2018), where the author investigates whether the trading volume in Bitcoin as well as its returns and realized daily volatility (RV) motivate and affect Bitcoin attention, measured by the SVI Google search volume index[6]. The author in Urquhart (2018) applies Vector AutoRegressive techniques and evidences, as a by product, that the log SVI index does not affect significantly Bitcoin log-returns and realized volatility for any of the three time series under investigation. Our results are consistent with above findings concerning the whole time series and the first period, while in the second subsample[7] we find a positive dependence between the log-SVI (X_1 in our notation) and Bitcoin returns; however the methodologies and the models fitted in this paper are much different from those in Urquhart (2018) and a clear-cut comparison is not feasible. In this paper we have not addressed any applications of the findings; next step in our research will be devoted to assess forecasting properties of the above selected models and to apply forecasting performance as a tool for model selection.

Acknowledgement(s)

This paper was written during the stay of G.F.T. at the Department of Finance and Risk Engineering, Tandon School of Engineering, New York University. She gratefully acknowledges financial support from Bank of Italy and Fondazione Cassa di Risparmio Perugia.

6. References

References

Almudhaf, Fahad. 2018. "Pricing efficiency of Bitcoin Trusts." *Applied Economics Letters* 25 (7): 504–508.

Blau, Benjamin M. 2017. "Price dynamics and speculative trading in bitcoin." *Research in International Business and Finance* 41 (Supplement C): 493 – 499. `http://www.sciencedirect.com/science/article/pii/S0275531917303057`.

Bukovina, Jaroslav, and Matúš Martiček. 2016. *Sentiment and Bitcoin volatility.* Technical Report. Mendel University in Brno, Faculty of Business and Economics.

Ciaian, Pavel, Miroslava Rajcaniova, and d'Artis Kancs. 2016. "The economics of BitCoin price formation." *Applied Economics* 48 (19): 1799–1815.

Cretarola, Alessandra, Gianna Figà-Talamanca, and Marco Patacca. 2017. "A sentiment-based model for the BitCoin: theory, estimation and option pricing." *SSRN Eletronic Journal* .

Da, Zhi, Joseph Engelberg, and Pengjie Gao. 2011. "In search of attention." *The Journal of Finance* 66 (5): 1461–1499.

[6]All variables are considered in the logarithmic scale

[7]The subsamples in Urquhart (2018) largely overlap the two periods in our analysis so we discuss them as if they were exactly the same

Dyhrberg, Anne Haubo. 2016. "Bitcoin, gold and the dollar - A GARCH volatility analysis." *Finance Research Letters* 16 (Supplement C): 85 – 92. http://www.sciencedirect.com/science/article/pii/S1544612315001038.

Engle, Robert Franklin. 1982. "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica: Journal of the Econometric Society* 50: 987 – 1007.

Katsiampa, Paraskevi. 2017. "Volatility estimation for Bitcoin: A comparison of GARCH models." *Economics Letters* 158 (Supplement C): 3 – 6. http://www.sciencedirect.com/science/article/pii/S0165176517302501.

Kristoufek, Ladislav. 2013. "BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era." *Scientific Reports* 3.

Kristoufek, Ladislav. 2015. "What Are the Main Drivers of the Bitcoin Price? Evidence from Wavelet Coherence Analysis." *PLoS ONE* 10 (4): e0123923.

Ljung, G. M., and G. E. P. Box. 1978. "On a measure of lack of fit in time series models." *Biometrika* 65 (2): 297–303. +http://dx.doi.org/10.1093/biomet/65.2.297.

Llorente, Guillermo, Roni Michaely, Gideon Saar, and Jiang Wang. 2002. "Dynamic Volume-Return Relation of Individual Stocks." *The Review of Financial Studies* 15 (4): 1005–1047. +http://dx.doi.org/10.1093/rfs/15.4.1005.

Nadarajah, Saralees, and Jeffrey Chu. 2017. "On the inefficiency of Bitcoin." *Economics Letters* 150 (Supplement C): 6 – 9. http://www.sciencedirect.com/science/article/pii/S0165176516304426.

Urquhart, Andrew. 2016. "The inefficiency of Bitcoin." *Economics Letters* 148 (Supplement C): 80 – 82. http://www.sciencedirect.com/science/article/pii/S0165176516303640.

Urquhart, Andrew. 2018. "What causes the attention of Bitcoin?" *Economics Letters* 166: 40 – 44. http://www.sciencedirect.com/science/article/pii/S016517651830065X.

Yelowitz, Aaron, and Matthew Wilson. 2015. "Characteristics of Bitcoin users: an analysis of Google search data." *Applied Economics Letters* 22 (13): 1030–1036.